ESTABLISHING SUPER-RESOLUTION IMAGING OF BIOSILICA-EMBEDDED PROTEINS IN DIATOMS

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

Philip Gröger

geboren am 23. Juli 1986 in Göttingen, Deutschland

Eingereicht am 12.05.2017

Die Dissertation wurde in der Zeit von Mai 2014 bis Mai 2017 im B CUBE - Center for Molecular Bioengineering der Technische Universität Dresden angefertig.

Gutachter: Prof. Dr. Michael Schlierf
Prof. Dr. Stefan Diez

If you want to know how nature works, we looked at it, carefully. Looking at it, that's the way it looks. You don't like it? Go somewhere else, to another universe where the rules are simpler, philosophically more pleasing, more psychologically easy.

- Richard Feynman

ABSTRACT

Diatoms feature the unique ability to form nanopatterned hierarchical silica cell walls with unprecedented accuracy and reproducibility. Gathering a deeper understanding of this process that is known as "biomineralization" is vitally important not only in the field of diatom research. In fact, the nanopatterning can also be exploited in the fields of material sciences or nanobiotechnology. According to the current understanding, the self-assembly of protein patterns along which biosilica is formed is key to this nanopatterning. Thus, in order to unravel the function of individual proteins that are involved in this biomineralization process, their structural organization has to be deciphered and correlated to morphological cell wall features that are in the order of tens of nanometer. Electron microscopy is able to resolve these features but does not provide protein-specific information. Therefore, a technique has to be established that is able to localize individual biosilica-associated proteins with nanometer precision.

To achieve this objective, single-molecule localization microscopy (SMLM) for the diatom *Thalassiosira pseudonana* has been pioneered and exploited to localize different biosilica associated proteins inside silica and after silica removal. To obtain quantitative data, an analysis workflow was developed including graphical user interfaces and scripts for SMLM visualization, clustering, and co-localization.

In order to find optimal labels for SMLM to target biosilica-embedded proteins, a comprehensive screening of photo-controllable fluorescent proteins has been carried out. Only three of six candidates were functional when embedded inside biosilica and fused to Silaffin3 – a protein that is tightly associated with the biosilica cell wall. Silaffin3 could be localized using SMLM with a localization precision of 25 nm. This allowed to resolve its structural organization and therefore identified Silaffin3 as a major component in the basal chamber of the fultoportulae.

Additionally, co-localization studies on cingulins – a protein family hypothesized to be involved in silica formation – have been performed to decipher their pattern-function relationship. Towards this end, novel imaging strategies, co-localization calculations and pattern quantifications have been established. With the help of these results, the spatial arrangement of cingulins W2 and Y2 could be compared with unprecedented resolution.

In summary, this work has laid ground for quantitative SMLM studies of proteins in diatoms in general and contributed insights into the spatial organization of proteins involved in biomineralization in the diatom *T. pseudonana*.

ZUSAMMENFASSUNG

Kieselalgen – auch Diatomeen genannt – verfügen über die einzigartige Fähigkeit, nanostrukturierte, hierarchisch aufgebaute Zellwände aus Siliziumdioxid – auch als Biosilica bekannt – mit beispielloser Genauigkeit und Reproduzierbarkeit zu bilden. Ein tieferes Verständnis für diesen Prozess, der als “Biomineralisation“ bekannt ist, ist nicht nur auf dem Gebiet der Grundlagenforschung zu Kieselalgen sehr bedeutsam, sondern auch für die Nutzung dieser Nanostrukturierung in den Materialwissenschaften oder der Nanobiotechnologie. Nach dem derzeitigem Stand der Wissenschaft wird diese Strukturierung durch die Selbstorganisation von Proteinmustern, an denen sich das Siliziumdioxid bildet, erreicht. Um die Funktion und das Zusammenspiel einzelner Proteine, die an diesem Biomineralisationsprozess beteiligt sind, entschlüsseln zu können, ist es essentiell ihre strukturelle Organisation aufzuklären und diese mit den morphologischen Zellwandmerkmalen zu korrelieren. Die Größenordnung dieser Merkmale ist im Bereich von Nanometern angesiedelt. Mit Hilfe der Elektronenmikroskopie können diese Biosilicastrukturen aufgelöst werden, jedoch ist keine proteinspezifische Information verfügbar. Ziel dieser Arbeit war es daher, eine Technik zu etablieren, die in der Lage ist, einzelne Biosilica-assozierte Proteine mit Nanometer-Präzision zu lokalisieren.

Um dieses Ziel zu erreichen, wurde Einzelmoleküllokalisationsmikroskopie (single-molecule localization microscopy, kurz: SMLM) beispielhaft in der Kieselalge *Thalassiosira pseudonana* etabliert. Die Position verschiedener Biosilica-assoziierte Proteine innerhalb des Biosilicas und nach dessen chemischer Auflösung wurde mit einer hohen räumlichen Auflösung bestimmt. Um quantitative Ergebnisse zu erhalten, wurde ein Analyse-Workflow entwickelt, der grafische Benutzeroberflächen und Skripte für die Visualisierung, das Clustering und die Kolokalisation von SMLM Daten beinhaltet.

Um optimale Markierungen für SMLM an Biosilica-eingebetteten Proteinen zu finden, wurde ein umfassendes Screening von photo-schaltbaren fluoreszierenden Proteinen durchgeführt. Diese wurden als Fusionsproteine mit Silaffin3, einem Protein, welches eng mit der Biosilica-Zellwand assoziiert ist, exprimiert. Es konnte gezeigt werden, dass nur drei von sechs Kandidaten funktional sind, wenn sie in Biosilica eingebettet sind. Silaffin3 konnte indirekt mittels SMLM mit einer Lokalisationsgenauigkeit von 25 nm detektiert werden. Dies erlaubte es, seine strukturelle Organisation aufzulösen und Silaffin3 als eine Hauptkomponente in der Basalkammer der Fultoportulae zu identifizieren.

Darüber hinaus wurden Kolokalisationsstudien an Cingulinen – einer Familie von Proteinen, die vermutlich an der Biosilica-Bildung beteiligt ist – durchgeführt, um die Beziehungen zwischen den Proteinmustern und deren Funktionen zu entschlüsseln. Zu diesem Zweck wurden neuartige Bildgebungsstrategien, Kolokalisationsberechnungen und Musterquantifizierungen etabliert. Mit Hilfe dieser Ergebnisse konnte die räumliche Anordnung der Cinguline W2 und Y2 mit einer beispiellosen Auflösung verglichen und eine mögliche Funktion in der Strukturierung postuliert werden.

Zusammenfassend hat diese Arbeit die Grundlagen für quantitative SMLM-Untersuchungen von Proteinen in Kieselalgen im Allgemeinen geschaffen und wertvolle Einblicke in die räumliche Organisation von Proteinen ermöglicht, welche an der Biomineralisation in *T. pseudonana* beteiligt sind.

CONTENTS

1 INTRODUCTION 1

1.1 Diatoms – a model system for biomineralization 3

1.2 Imaging of biosilica and associated organic components 8

1.3 Single-molecule localization microscopy (SMLM) 10

2 METHODS & METHOD DEVELOPMENT FOR SMLM DATASETS 17

2.1 Super-resolution reconstruction 19

2.2 Tools for SMLM resolution estimates 21

2.3 Voronoi tessellation for noise-removal and cluster estimation 25

2.4 Tools for SMLM cluster analysis 27

2.5 Coordinate-based co-localization 32

2.6 PairRice – A novel algorithm to extract distances between cluster pairs 33

2.7 SiMoNa – A new GUI for exploring SMLM datasets 35

3 RESOLUTION OF THE SMLM SETUP TESTED WITH DNA ORIGAMI NANOSTRUCTURES 41

3.1 DNA origami as a length standard 42

3.2 Global resolution estimates 44

3.3 Local resolution estimates 47

3.4 Conclusion 53

4 EVALUATION OF PHOTO-CONTROLLABLE FLUORESCENT PROTEINS FOR PALM IN DIATOMS 55

4.1 Selecting PCFPs to minimize interference with the diatom autofluorescence. 56

4.2 Screening results for cytosolic and biosilica-embedded PCFPs 58

4.3 The underlying conversion mechanism 61

4.4 Conclusion 63

5 IMAGING THE SIL3 MESHWORK 65

5.1 Analyzing protein layer thickness using tpSil3-Dendra2 65

5.2 Imaging the valve region using tpSil3 68

5.3 Resolution and localization parameters of tpSil3 70

5.4 Conclusion 72

6 DECIPHERING CINGULIN PATTERNS WITH CO-LOCALIZATION STUDIES 73

6.1 A two-color cingulin construct for PALM-STORM 73

6.2 Steps towards PALM-STORM: screening, alignment, and imaging routine... 76

6.3 Co-localization studies: quantification, clustering, and correlations 83

6.4 Conclusion 91

7 OUTLOOK 93

8 MATERIALS & METHODS 97
8.1 Microscope specifications 97
8.2 DNA origami annealing and AFM measurements 99
8.3 Diatom sample preparations 100
8.4 Fluorescence imaging conditions 102
8.5 Buffer systems 103
9 APPENDICES 105
9.1 Tables and Protocols 105
9.2 Satellite projects 112
9.2.1 Quantitative fluorescence intensity analysis of 3D time-lapse confocal microscopy data in diatoms 112
9.2.2 Applying neural networks to filter SMLM localizations 118
9.2.3 In vivo imaging at super-resolution conditions using SOFI 121
9.2.4 Quantifying chromatic aberrations in the microscope using fiducials ... 123
10 REFERENCES 127

LIST OF FIGURES

Figure 1.1 The morphological spectrum of diatoms 3

Figure 1.2 SEM image of isolated biosilica from *T. pseudonana* 4

Figure 1.3 The morphology of diatom *T. pseudonana* 5

Figure 1.4 Cingulin localization and microring imaging 8

Figure 1.5 Commonly used imaging techniques and their resolution 10

Figure 1.6 The SMLM imaging process 12

Figure 1.7 Size comparison of fluorescent probes and linker 13

Figure 2.1 Drift correction procedure based on cross-correlations 20

Figure 2.2 Fitting single emitters with a 2D Gaussian 21

Figure 2.3 The effect of localization density and precision 23

Figure 2.4 Fourier-ring correlation workflow 24

Figure 2.5 Overview of the Voronoi workflow...27

Figure 2.6 Ripley's K function used to analyze spatial point patterns..29

Figure 2.7 Pairwise distance fitting scheme..30

Figure 2.8 Coordinate-based co-localization for simulated data..33

Figure 2.9 SiMoNa main window...36

Figure 2.10 Spotting faulty localizations with SiMoNa 38

Figure 2.11 Comparison of different super-resolution reconstruction algorithms 39

Figure 3.1 Sample design and AFM images of the DNA origami 43

Figure 3.2 STORM images of a 76 nm origami..44

Figure 3.3 Localization parameter for the 76 nm origami...46

Figure 3.4 Nearest neighbor and Fourier ring correlation...47

Figure 3.5 Clustering determination via Ripley’s H function 47

Figure 3.6 Comparison of widefield and super-resolution image for a 76 nm origami 48

Figure 3.7 Length measurements in a 76 nm origami 50

Figure 3.8 Evaluation of different fitting methods for different origami samples 51

Figure 3.9 Evaluation of different fitting methods for simulated data 52

Figure 4.1 Emission spectra of selected PCFPs and live *T. pseudonana* 58

Figure 4.2 Screening of the fluorescent proteins and their activation capability 59

Figure 4.3 Photo-conversion statistics of active tpSil3 clones 61

Figure 4.4 Sequence similarities for PCFPs..62

Figure 4.5 Types of chromophore conversion...63

Figure 5.1 tpSil3 illumination scheme 66

Figure 5.2 tpSil3-Dendra2 thickness measurements 68

Figure 5.3 PALM of tpSil3 in the valve region...69

Figure 5.4 SEM image of the inside of isolated valve biosilica of *T. pseudonana* 70
Figure 5.5 Localization parameter for tpSil3 fusion proteins 71
Figure 6.1 Transformation and imaging preparation workflow for the two-color cingulin construct .. 74
Figure 6.2 FCS measurements of the GFPTrap with eGFP and Dendra2 76
Figure 6.3 Widefield images of live diatoms screened for eGFP and Dendra2 fluorescence 77
Figure 6.4 Widefield images of CinW2 & CinY2 tagged microrings 78
Figure 6.5 Super-resolution images of CinW2 & CinY2 tagged microrings 79
Figure 6.6 Channel clean-up and alignment workflow 80
Figure 6.7 Super-resolution scatter plot of microrings before and after filtering and alignment 81
Figure 6.8 Super-resolution scatter plot details of microrings 82
Figure 6.9 Localization precision and nearest neighbor for both channels 83
Figure 6.10 Co-localization and cluster analysis workflow 84
Figure 6.11 Influence of CBC parameter 85
Figure 6.12 Co-localization scatter plots 86
Figure 6.13 Co-localization histograms 87
Figure 6.14 DBSCAN for cingulin cluster identification 88
Figure 6.15 Defining the cluster area and cluster ellipticity 89
Figure 6.16 Area and shape statistics of the detected clusters 90
Figure 6.17 From two-color clusters to cluster overlap 90
Figure 6.18 Cingulin cluster parameter overview and model hypothesis 92
Figure 7.1 Proposal for a three-color cingulin construct 94
Figure 8.1 Vector design for cytosolic and Sil3 expression 101
Figure 9.1 Image acquisition illustrations and time-lapse fluorescence images 113
Figure 9.2 Image registration to correct sample drift 114
Figure 9.3 Fluorescence quantification using rectangular areas 115
Figure 9.4 Geometric definitions for the correction factor calculation 117
Figure 9.5 Network performance for CinW2-Dendra2 cingulin data filtering 120
Figure 9.6 SOFI on tpSil3-Dendra2 in live cells 122
Figure 9.7 Quantifying chromatic aberrations with fluorescent beads 125

LIST OF TABLES

Table 1 Organic components of the *T. pseudonana* frustule 6
Table 2 Overview of clustering algorithms 28
Table 3 Overview of PCFPs 57
Table 4 Cluster and overlap statistics 91

LIST OF ABBREVIATIONS

Microscopy techniques

AFM	atomic force microscopy
PALM	photo-activated localization microscopy
SEM/TEM	scanning/transmission electron microscopy
SIM	structured illumination microscopy
SMLM	single-molecule localization microscopy
SOFI	super-resolution optical fluctuation imaging
STED	stimulated emission depletion (microscopy)
STORM	stochastic optical reconstruction microscopy
ThreeB/3B	Bayesian analysis of blinking and bleaching

Analysis-related

CBC	coordinate-based co-localization
FRC	Fourier ring correlation
GMD	expectation maximization algorithm for a Gaussian mixture distribution
GUI	graphical user interface
MATLAB	“matrix laboratory”, a numerical computing environment and programming language
KMC	k-means clustering
PSF	point-spread function

Biology-related

bp	base pairs
(e)GFP	(enhanced) green fluorescent protein
FP(s)	fluorescent protein(s)
MQ water	Milli-Q®, ultrapure water
PCFP(s)	photo-controllable fluorescent protein(s)
SDV(s)	silica deposition vesicle(s)
T. pseudonana	*Thalassiosira pseudonana*

1 INTRODUCTION

“If there is magic on this planet, it is contained in water”, said anthropologist Loren Eiseley in 1957 in his famous book “The Immense Journey”. Although some of this magic in the aquatic kingdom is unraveled, it still is a place of the great unknown.

Diatoms are definitely among the mysterious and little-known species of the aquatic environment that deserve more profound scientific exploration due to their immense ecological importance as well as due to their manifold possible applications in nanotechnology. These diverse microscopic algae include up to 100’000 species [1] and generate 45% of the organic matter that serves as the first link of the food chain for the life in the sea [2]. This equals one quarter of the global primary production [3]. Additionally, about 20% of photosynthesis on earth – as much as all of earth’s rain forests – is performed by diatoms [4]. Besides the ecological role of diatoms, they are masters of a fascinating capability that has been the main motivation to perform this doctoral work (and gives rise to their characteristic aesthetic look): biomineralization.

Biomineralization is the molecular process of forming inorganic material by biological systems and is widespread in all kingdoms of life. A huge variety of inorganic material (e.g., silica, calcite, magnetite) is produced by different organisms that serve different functions and tasks (e.g. stability through skeleton formation, protective shells, teeth, and even sensors). Scientists – especially material scientists – have been interested to understand the underlying biochemical principles of biomineralization and to translate them into technological applications. In the search for a model system to study biomineralization, diatoms – who encase themselves in a silica cell wall – are a promising choice as their capability to reproducibly form hierarchical, regular pore patterns out of biosilica is unprecedented and has fascinated scientist since the advent of microscopy (see Figure 1.1 for historic diatom drawings from 1904 by Ernst Haeckel). Additionally, culturing protocols are available, different genetic transformation systems have been established for a few species [5] and the full genome of two species has been published recently [6], [7]. Thus, important prerequisites to understand biomineralization on the molecular level and to engineer genetically modified diatoms for specific tasks are given. One of those two fully sequenced species, *Thalassiosira pseudonana*, serves as a model system for this doctoral work and will be introduced in chapter 1.1, where also key players of the silicon metabolism will be highlighted.

Diatom sizes range from 2 to 500 µm and their hierarchical biosilica pore pattern exhibit features sizes from 1 µm down to tens of nanometer. Therefore, the availability of a microscopy technique that can visualize these intricate patterns is key to infer a connection to the underlying biomolecule meshwork. As conventional light microscopy cannot resolve features under 200 nm other microscopy techniques have to be used in complement. Chapter 1.2 will present the most frequently used imaging techniques for biomineralization studies and motivate the need for a technique that is able to visualize biomolecule-specific patterns at a high resolution while being *in vivo* compatible. The technique of choice is single-molecule localization microscopy which is introduced and explained in more detail in chapter 1.3.

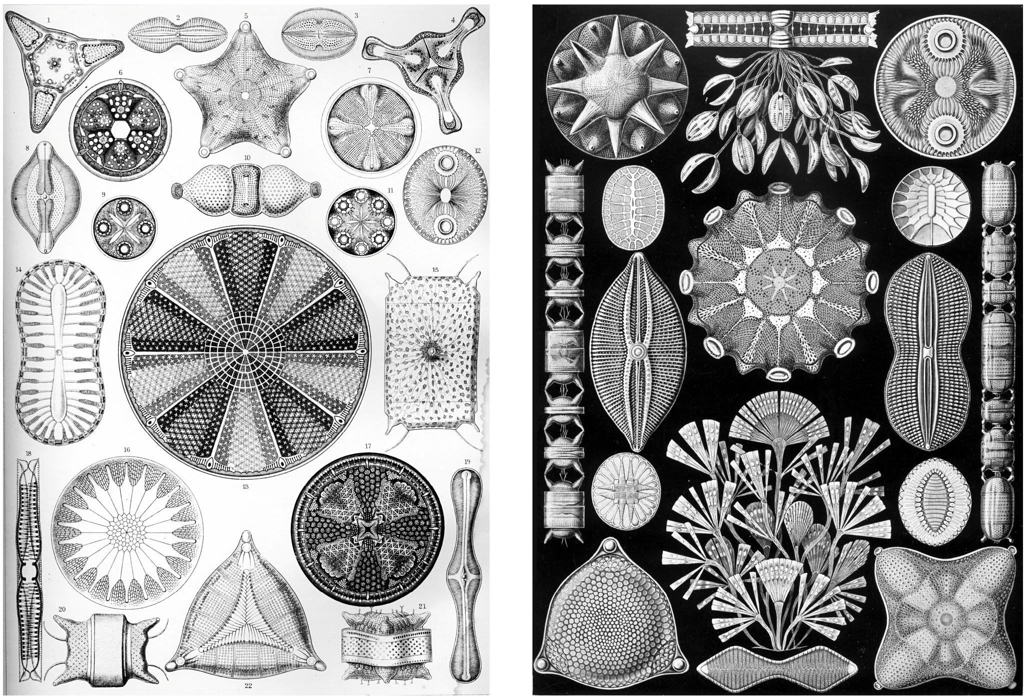


Figure 1.1 The morphological spectrum of diatoms Plate 4 and 84 of Ernst Haeckel's "Kunstformen der Natur" from 1904 displaying a wide range of diatom illustrations [8].

1.1 Diatoms – a model system for biomineralization

Diatoms are unicellular eukaryotic phototropic algae that occur abundantly in nearly every aquatic habitat. One of their defining features is the siliceous cell wall – the frustule. Figure 1.1 displays the immense morphological diversity of diatoms and already indicates their capability to form patterned structures on different length scales [8].

The model organism Thalassiosira pseudonana. *Thalassiosira pseudonana* (*T. pseudonana*) is a marine centric diatom. It has been the first eukaryotic marine phytoplankton whose 34 mega-base pair long genome has been fully sequenced [6]. This opened up many possibilities for research in the context of functional genomics and made *T. pseudonana* a most suitable model organism to study biomineralization. Its silicified cell wall (frustule) has a cylindrical shape with a diameter and height of around 5 µm (shown in Figure 1.2 using SEM). *T. pseudonana* features a very elaborate hierarchical, nano-patterned frustule (Figure 1.3A, B and D). The cell wall is divided into two planar valves and a lateral surface area consisting of circular girdle-bands in between the valves (Figure 1.3A). The valves contain a central as well as multiple marginal fultoportulae. These are tubular protrusions through the valve found in many centric diatoms [9]. Additionally the valves exhibit a fine ridge network radiating from the center with nanopores in between (average diameter 18 nm) [10]. The chloroplasts of *T. pseudonana* are easily visible in fluorescence microscopic images as oblate objects featuring strong autofluorescence under red illumination (647 nm) (Figure 1.3C). As the frustule is a rigid structure, cell growth and division can only be accomplished by synthesis of new biosilica. For *T. pseudonana*, cell growth along the cylinder axis is enabled through the synthesis of new girdle bands. Cell division on the other hand is facilitated by synthesis of two new valves in the valvar plane (i.e. the girdle equator) that are exocytosed upon maturation and ultimately split the cell into two daughter cells [11].

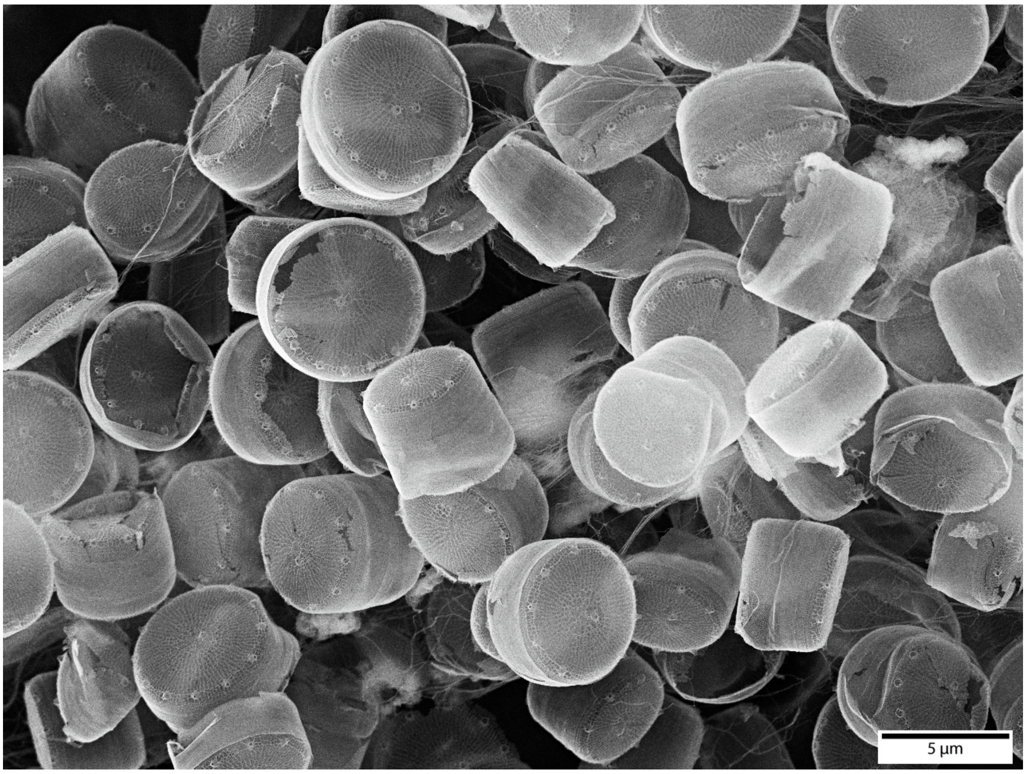


Figure 1.2 SEM image of isolated biosilica from *T. pseudonana* Agglomeration of dry biosilica cell walls from *T. pseudonana*. The reproducible cell wall architecture can be seen. Several diatoms appear damaged due to vortexing, centrifugation, dehydration before imaging, or vacuuming during imaging.

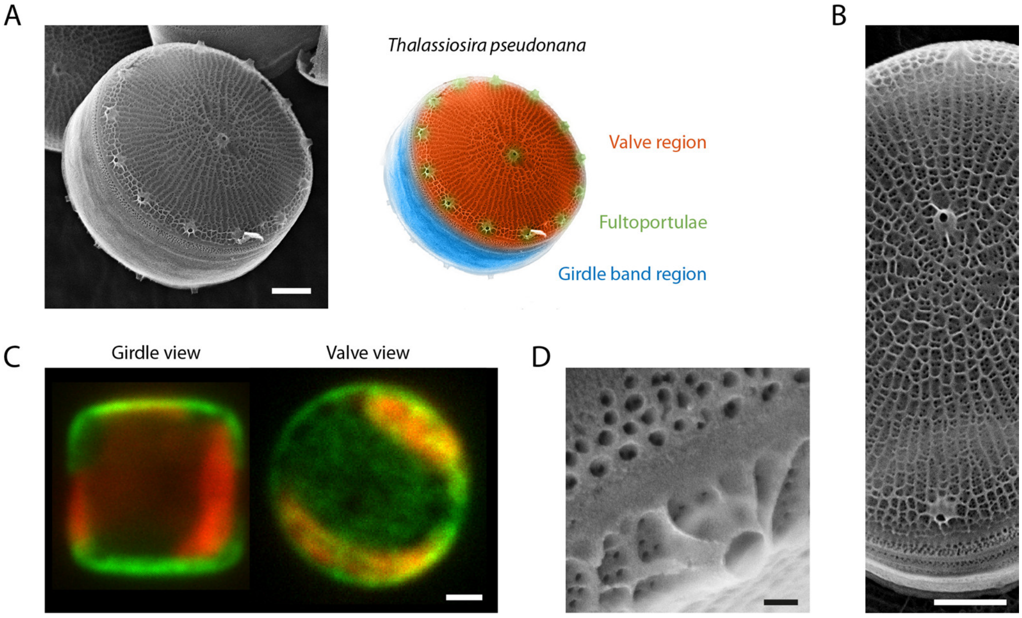


Figure 1.3 The morphology of diatom *T. pseudonana* A) Left: SEM image of the *T. pseudonana* biosilica displaying parts of the girdle band region and the planar valve region featuring the prominent fultoportulae and a ridge network. Right: False coloring of the EM picture to highlight structural features of the *T. pseudonana* biosilica. B) SEM image showing the valve and its intricate biosilica pattern in more detail. C) *In vivo* confocal microscopy images of two *T. pseudonana* cells with the cell wall protein Sil3 fluorescently tagged with GFP in green (488 nm excitation) in 'girdle view' (left) and 'valve view' (right). The red fluorescence is caused by autofluorescence of the chloroplasts (647 nm excitation). All scale bars are 1 µm. D) SEM image of a fultoportulae close-up with its surrounding pores. Scale bar 100 nm.

The biomineralization process. TEM studies revealed that biomineralization in diatoms takes place in so called silica deposition vesicles (SDVs) which are specialized intracellular compartments [12]. The SDVs host the silica morphogenesis process and deposit the silica valves or girdle bands on the cell surface by exocytosis. Various organic molecules have been identified that are associated with the biosilica and play a role in biosilica formation. However, still little is known about their biological function, interplay and intracellular assembly.

Organic components of the frustule. Besides the inorganic silica, multiple types of biomolecules have been found inside the *T. pseudonana* frustules [11] (Table 1): On the one hand there are components that become soluble upon dissolving the biosilica which is achieved by treatment with ammonium fluoride. This includes the silaffin proteins [13], [14], silacidin phosphopeptides [15], as well as long-chain polyamines [16]. The diverse group of long-chain polyamines (featuring complex methylation patterns) seem to be an important component in pattern formation during biomineralization. On the other hand, insoluble components of the organic material have been discovered that form organic matrices after biosilica dissolution. Among those the protein families of cingulins [17] and silicanins [18] as well as polysaccharides [19] were identified so far. The polysaccharides form a chitin-based meshwork that is associated with parts of the biosilica. Additionally, protein-based microrings and microplates remain (Figure 1.4B). These organic microrings and microplates are associated with the girdle band and valve biosilica, respectively. As silaffins and cingulins will be studied further in this doctoral work, they will be introduced explicitly in the following paragraphs.

Table 1 Organic components of the *T. pseudonana* frustule Overview of all known organic components associated with the biosilica cell wall with citations of their original discovery. The last column highlights the solubility after the biosilica has been dissolved by ammonium fluoride treatment as applied in this work and detailed in the methods chapter 8.3.

Organic compound	Type	Members	Soluble after biosilica dissolution
Silaffin [13], [14]	Protein	sil1/2L, sil1/2H, sil3, sil4	Yes
Silacidin [15]	Peptide	silacidin A, B & C	Yes
Cingulin [17]	Protein	CinW1-W3, CinY1-Y3	No
Silicanin [18]	Protein	SiMat1-7	No
Long-Chain Polyamine [16]	Polyamine	3 classes (propylenediamine, pturescine, sperimidine basis)	Yes
Chitin [19]	Polysaccharide	-	No

Silaffins. Six polycationic peptides have been identified that are associated with biosilica but are soluble by ammonium fluoride treatment: the silaffins (after “proteins with silica affinity”) tpSil1H -1L, -2H, -2L, -3 and -4. Here, “H” and “L” denote the heavy and light form, before and after proteolytic processing, respectively. Biochemical analyses revealed that these silaffins are heavily phosphorylated, are composed of (often repetitive) serine and lysine motives, and contain several post-translational modifications. Additionally, they contain at least a single pentalysine cluster which has been shown to mediate silica targeting [20]. So far, their spatial arrangement has only been investigated with conventional fluorescence microscopy using GFP fusion proteins. The results of these studies suggest that tpSil1 and tpSil2 are located exclusively in the girdle band region in circular patterns around the frustules [20]. tpSil3 on the other hand is located in all parts of the cell wall [21]. Studies localizing this proteins below the diffraction limit have not been carries out, yet. This makes tpSil3 an ideal target protein to establish single-molecule localization microscopy, as will be shown in chapter 4 and 5.

Cingulins. Cingulins are a conserved class of proteins with highly repetitive domain structure that were identified by Scheffel *et al.* in a bioinformatics screen in *T. pseudonana* [17]. Their amino acid composition features silaffin-like regions that are rich in serine and lysine residues. Currently, 2 families with 3 members each have been identified that are additionally rich in tyrosine (CinY1-3) or tryptophan (CinW1-3). GFP fusion proteins have been generated and could map all cingulins exclusively to the girdle-band region of the biosilica of *T. pseudonana* (Figure 1.4A). It is conceivable that cingulins can induce silica precipitation due to their silaffin-like domains [18]. They are associated with biosilica and remain insoluble even after silica dissolution. This insoluble material has been identified as part of the silica-forming microrings. In electron micrographs, the organic filaments resemble the pore-free regions of the girdle bands (Figure 1.4B&C). Therefore it can be hypothesized that cingulins are involved in silica morphogenesis. The spatial relationship and thus the functional interplay of individual cingulins with each other is still unknown. They could be distributed equally within the organic filament or each cingulin could fulfill a specific task at a specific location within the microring. Conventional fluorescence microscopy cannot resolve the fine features of the microrings and therefore is not sufficient to visualize differences in cingulin distributions within a single microring. To overcome this limitation and to answer these questions was the main motivation to study these proteins with super-resolution microscopy (chapter 6).

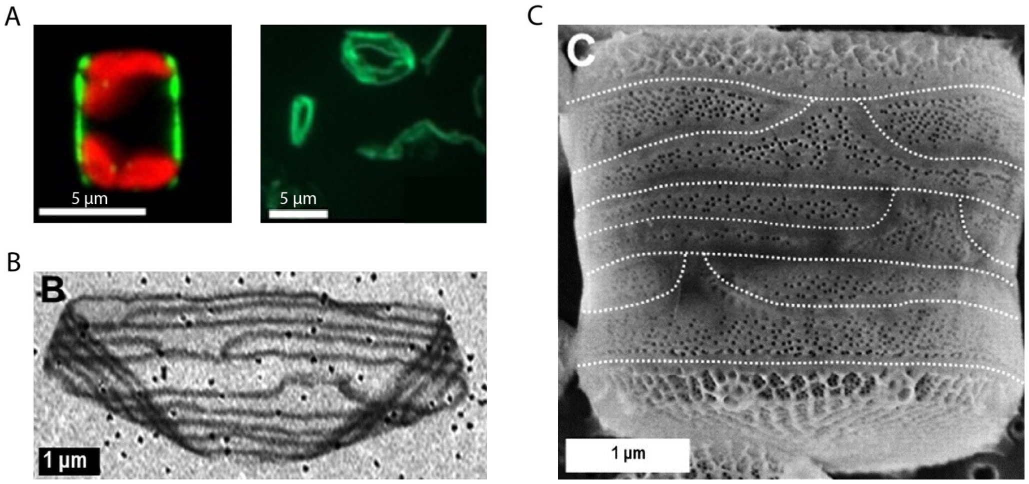


Figure 1.4 Cingulin localization and microring imaging A) Confocal fluorescence microscopy images of CinW2-GFP in live cells (left) and after complete demineralization (right). The red fluorescence is caused by chloroplast autofluorescence. B) SEM image of an individual microring displaying circular organic filaments. C) SEM image of intact biosilica of an individual T. pseudonana cell. The white dotted line delineate the pore-free regions of the girdle bands. All images and figure legend adopted from [17].

1.2 Imaging of biosilica and associated organic components

Several approaches are conceivable to study the aforementioned organic components of the biosilica cell wall to ultimately understand their functional interaction. While molecular genetic, biochemical, proteomic, or genomic analysis of individual compounds is possible *in vitro* and ultimately led to the discovery of the biomolecules listed in Table 1, imaging enables *in situ* studies and therefore acquires spatial information. This spatial information is of pivotal importance to understand the process of morphogenesis. The most prominent imaging techniques that are available to study organic compounds related to biomineralization will be introduced in this chapter and the need to establish an *in vivo* compatible super-resolution technique for this field is motivated. An overview of the discussed imaging techniques and their resolution capacities can be found in Figure 1.5.

SEM/TEM. Extensive studies using transmission electron microscopy have been performed to elucidate the biomineralization process inside the diatom cell, with the visualization of the SDVs just being one important discovery [22]. Scanning electron microscopy could be used efficiently to probe the diatoms nano-patterned biosilica cell wall morphology to reveal structural features and insights into the biosilification process [23]. Naturally, these studies can only visualize static snapshots of cryo-fixed (TEM) or dehydrated (SEM) cells in a vacuum chamber. Biomolecule-specific information can principally be obtained by using Immunogold staining [24]. However, this technique is difficult to apply for a majority of organic compounds of the *T. pseudonana* frustule, as antibodies for insoluble proteins are hard to synthesize.

AFM. Atomic force microscopy has been used to image the microrings of *T. pseudonana* and their mineralization capabilities [17]. Although the spatial resolution of AFM is in the nanometer range, it can only be used to probe the surface topography.

Fluorescence microscopy. For *in vivo* imaging of diatoms, conventional light microscopy has been the exclusive tool to date and even dates back to the Victorian era, where diatoms were one of the most collected microscopic object. The invention fluorescence microscopy allowed insight into cellular process throughout the cell cycle by using a broad palette of dyes and stains. With the sequencing of the genome of *T. pseudonana* [6] and the establishing of genetic transformation systems [5], [25], [26], the basis has been laid to label proteins with fusion constructs using genetically encoded fluorescent probes like GFP and to locate them within the cell. Since spatial resolution is still limited by the diffraction limit of light (as will be elaborated in the following chapter 1.3) to around 160 nm, there is a huge information gap between "high-resolution" biosilica morphology information from SEM studies and the protein-specific "low-resolution" information from fluorescence microscopy.

Single-molecule localization microscopy. Single-molecule localization microscopy (SMLM) is able to bridge this gap by offering a resolution down to tens of nanometers while still preserving the advantages of fluorescence microscopy like e.g. the possibility of *in vivo* imaging and specific labeling. While the technique itself will be described in detail in the following chapter 1.3, Figure 1.5 already emphasizes SMLM's position between classical confocal fluorescence microscopy and high-resolution electron microscopy. Typical feature sizes of the *T. pseudonana* frustule are given, showing e.g. that pores and fultoportulae with their sizes of roughly 10 and 100 nm, respectively, cannot be resolved by widefield microscopy.

Ensemble-based super-resolution techniques. Two further fluorescence- and ensemble-based super-resolution techniques that are worth mentioning are structured illumination microscopy (SIM) [27] and stimulated emission depletion (STED) [28]. As both techniques have not been applied in this thesis, they will be listed in Figure 1.5 for the sake of completeness but will only explained very briefly here. SIM illuminates the sample with striped patterns of light to generate interference patterns. These can be analyzed to reconstruct the original image at an increased lateral resolution of around 100 nm. STED can be applied using a modified scanning confocal microscope, where an additional doughnut-shaped depletion beam suppresses fluorescence emission from fluorophores located off the center of the excitation. Thereby the lateral resolution can be increased to around 30 nm. Although both techniques generally do not reach the spatial resolution of SMLM, they might find application in diatom imaging e.g. if a high temporal resolution becomes necessary.

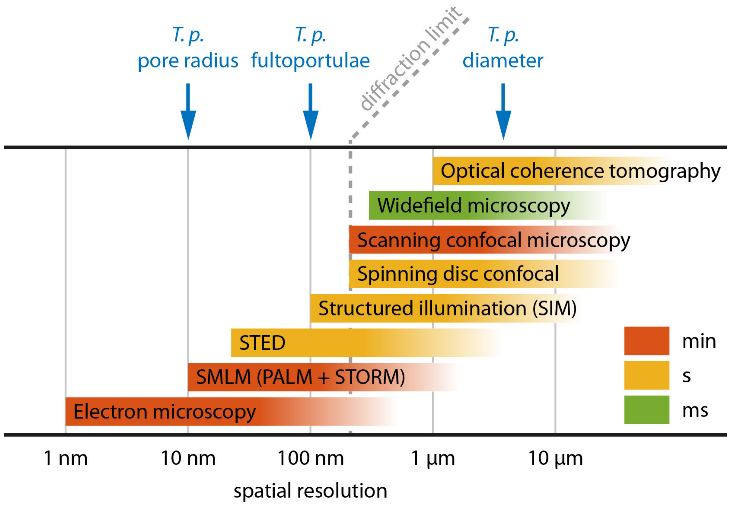


Figure 1.5 Commonly used imaging techniques and their resolution Overview of different imaging techniques used in biology and their resolution ranges. Colored by a typical "fast" acquisition times for one image. Typical feature sizes of *T. pseudonana* (*T. p.*) are noted above for comparison.

1.3 Single-molecule localization microscopy (SMLM)

Before elaborating the technical details and image acquisition routine of SMLM, it is essential to consider the field of bio-imaging before the rise of super-resolution techniques in order to understand the importance of this technique for the scientific community.

The Abbe limit. For lens-based microscopy, the maximum resolution is confined by the diffraction limit of light as defined by Ernst Abbe in 1873:

$$d = \frac{\lambda}{2 \cdot NA} \qquad (1)$$

It is based on the idea that two objects can only be discriminated and separated if they are separated by a distance d defined by the wavelength of light λ and the numerical aperture NA of the optical system. While lens design has advanced since

1873, even with the brightest objectives that feature an NA of around 1.5 and by using a low imaging wavelength of 500 nm, the resolution limit remains above 160 nm. With the novel super-resolution or sub-diffraction techniques featuring different imaging modalities and strategies finally a way to circumnavigate the Abbe limit has been discovered.

The advent of the single-molecule approach. With the invention of total internal reflection fluorescence (TIRF) microscopy in the early 1980s [29], a tool has become available to study fluorescence in a very thin and defined layer within hundreds of nanometer above the coverslip surface. By these means, it became easily possible to investigate single immobilized emitters by reducing the sample concentration. Although the image of a single fluorescent probe is still blurred by the refraction of light, its position can be determined very accurately by fitting the light distribution with a 2D-gaussian that represent an approximation of an airy disk and retrieving the center position. For this approach, the emitters have to be spaced sufficiently apart from each other so that the center position can be determined accurately and the number of photons emitted per single fluorophore per frame has to be sufficient to ensure a proper centroid determination. Given those two requirements, the localization precision Δx scales approximately with the inverse square root of total photons per peak N, and can thus be increased arbitrarily beyond the diffraction limit:

$$\langle(\Delta x)^2\rangle = \frac{\sigma^2}{N} \qquad (2)$$

The single molecules imaging approach. The idea to localize single molecules to reconstruct a complete image has already been used extensively to track individual particles or emitters for e.g. surface diffusion analysis [30]. Using this technique for fluorescence imaging – where millions of fluorescent probes emit simultaneously – is not possible *per se*. In 2006 however, multiple research groups came up with similar approaches to form a super-resolved image via single-molecule localizations – hence, the name single-molecule localization microscopy (SMLM). Synonyms include super-resolution pointillist imaging, coordinate-stochastic super-resolution microscopy or just nanoscopy. The individual techniques, termed photo-activated localization microscopy (PALM) [31], fluorescence photo-activation localization microscopy (FPALM) [32], and stochastic optical reconstruction microscopy (STORM) [33] all rely on the same workflow (see Figure 1.6):

1. The target of interest has to be tagged with a fluorescent probe (e.g. a photo-controllable fluorescent protein or a fluorescent dye)

2. The fluorescent probes have to reside or be transferred to an OFF-state
3. A very low percentage is stochastically transferred into an ON-state
4. In the ON-state they can be excited and imaged as individual emitters if their spatial separation is larger than 200 nm
5. The peak position of each probe in a single frame is determined with high precision and stored
6. After having been imaged, these few probes have to be transferred again into an OFF-state (for example by bleaching)
7. Then another subset of probes is transferred into the ON-state
8. This cycle is repeated thousands of times and all the localizations are pooled together
9. These localizations are used to reconstruct the final super-resolved image by plotting their coordinates into a single image

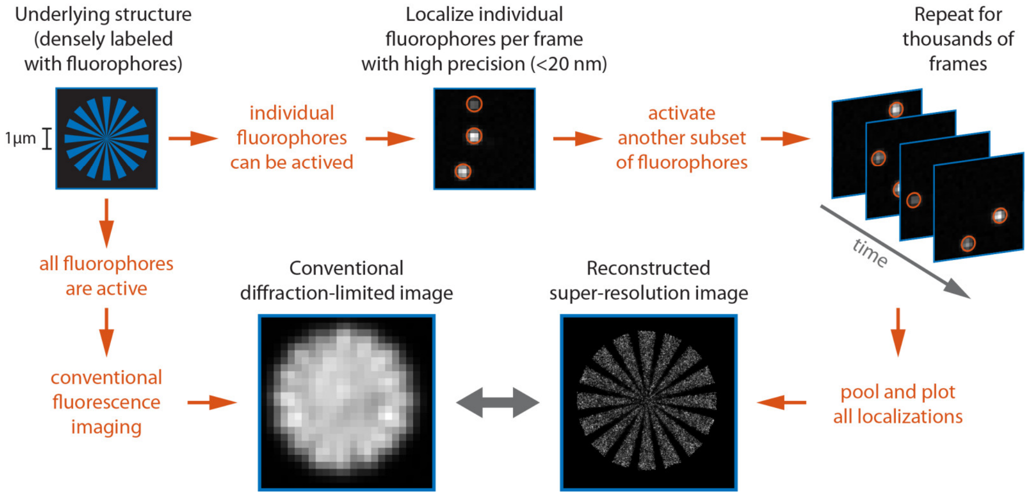


Figure 1.6 The SMLM imaging process The super-resolution workflow is illustrated for a Siemens-star which has a size of a few microns. The underlying structure is tagged with fluorescent markers of which a low percentage has to be switched on. Multiple frames are recorded and the position of individual fluorescent events is detected. These localizations are summed up and a super-revolved image is formed. By comparison with a conventional diffraction-limited image, the resolution gain is obvious. A detailed description of the workflow is presented in the main text.

Fluorescent probes for SMLM. The most important requirement of a fluorescent probe for SMLM is its ability to switch between a non-fluorescent OFF and fluorescent ON state. The availability of a mechanism to actively control the blinking behavior is preferable, however also pure stochastic blinking is sufficient. Some of the fluorescent probes that are available for standard widefield microscopy are also suited for SMLM (Figure 1.7). The two most commonly used probes for SMLM are fluorescent proteins and fluorescent dyes, correlating to the two SMLM techniques PALM and STORM, respectively. For PALM, classical fluorescent proteins (e.g. EGFP) have to be modified to be controllable by light (e.g. photo-activatable GFP - PA-GFP) [34], [35]. These so called photo-controllable fluorescent-proteins (PCFPs) can be switched OFF by one wavelength (normally the excitation wavelength) and switched ON by a second wavelength (most commonly by near-UV light). By genetically introducing a fusion protein of the FP (fluorescent protein) and the protein of interest the label can be placed very specifically. Thus, PALM shows its strength in noninvasive and *in vivo* studies of biological processes. For STORM on the other hand, fluorescent dyes that are linked to the protein of interest by established labeling techniques such as immunofluorescence. Blinking is induced using a buffer containing thiols (e.g. Cysteamine) that shift the fluorophore into a dark state from which it can be recovered using UV light or the excitation wavelength [36], [37]. As these fluorophores yield higher photon counts than fluorescent proteins, the localization precision and by extension the resulting image resolution is generally higher [38]. However, the use of STORM is limited for *in vivo* studies as an additional labeling step (e.g. via antibodies) is necessary and therefore, the labeling target has to be accessible. Apart from fluorescent proteins and fluorescent dyes, quantum dots and nitrogen-vacancy fluorescent nano-diamonds have been used in preliminary studies as SMLM probes and feature some promising properties, e.g. resistance to photobleaching [39], [40].

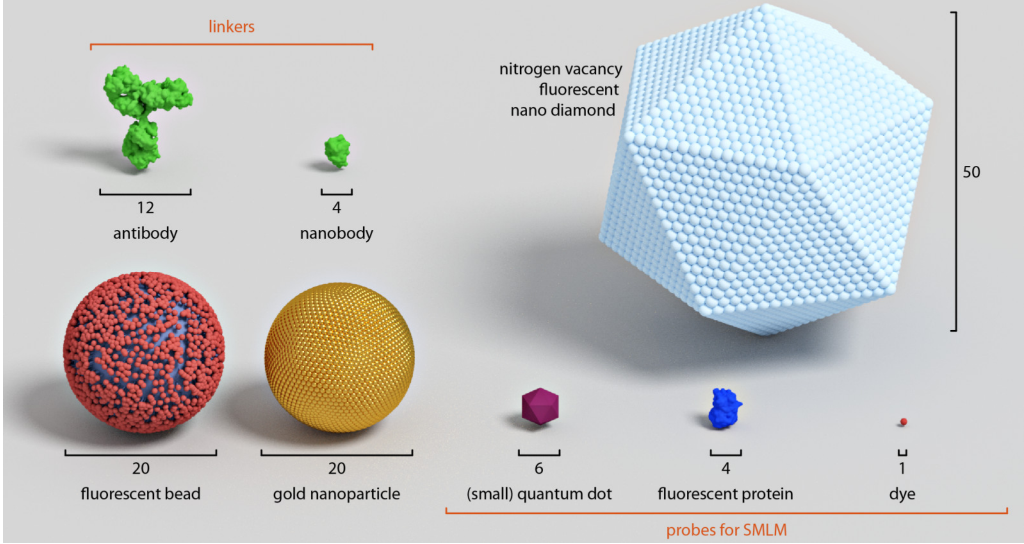


Figure 1.7 Size comparison of fluorescent probes and linker Qualitative size comparison of various fluorescent probes. An antibody (pdb:1IGT) and cameloid nanobody (pdb:1MEL) are shown in the top left as commonly used linker molecules. For the fluorescent protein, EGFP (pdb:1GFL) has been chosen. Length scales are in nm.

SMLM features some significant differences compared to traditional widefield fluorescence microscopy in terms of image acquisition, reconstruction, evaluation and analysis. The following paragraphs will introduce these specific features and highlight the most substantial strengths and limitations of SMLM.

Image resolution of SMLM. The unprecedented spatial resolution of SMLM as compared to conventional fluorescence microscopy is naturally its biggest strength. With novel microscope designs, this resolution could even be extended into the third dimension to allow volumetric 3D imaging [41]. On the downside, this high spatial resolution comes at the expense of temporal resolution. In order to form a dense super-resolution image, it is necessary to record thousands of frames which typically takes minutes. During these minutes, the tagged structure should not exhibit dynamics, and drift of the sample stage has to be minimized. In order to counter these limitations, efficient drift correction tools are available [42], [43] as well as methods to reduce the imaging time to seconds which even enabled visualization of slow dynamic processes with SMLM [44]. The quantification of image resolution in widefield fluorescence microscopy is easily defined by the point spread function of the imaging system. For SMLM, this term is not as straightforwardly tangible. The initial localization precision of each emitter does not necessarily reflect the final image resolution, as a sparse point distribution will introduce artificial gaps in the structure of interest. Yet, solutions to overcome this problem and to define a “real” resolution are available and will be discussed in more detail in chapter 2.2.

Compatibility of SMLM. In order to enable SMLM at a laser fluorescence microscope, some modifications are necessary. A sensitive camera (EMCCD or sCMOS) that can detect single fluorophores and a UV laser to specifically control photo-activation of the fluorescent probes are helpful. Apart from that, only changes to the sample preparation are necessary. For PALM, photo-controllable fluorescent proteins have to be used. Therefore, established protocols using fluorescent proteins can be used and *in vivo* measurements could be made possible. For STORM, labeling is performed using immunofluoresnce and the usage of a “blinking buffer”, mostly based on thiols that transfer the fluorescent dyes into the dark state, displays the only addition to conventional fluorescence imaging using these probes.

Image analysis of SMLM datasets. The single-molecule nature of SMLM adds multiple steps of image processing until a super-resolution image is formed (single-molecule localization and image reconstruction). These steps have impact on the appearance of the final image and thus have to be chosen carefully. Hence, chapter 2.1 will introduce the image reconstruction in detail. The single-molecule information, on the other hand, supplies the user with a variety of photo-physical parameters of the fluorescent probes additionally to their spatial coordinates. Moreover, image analysis is not prone to binning artifacts created by discrete pixel values as is the case in widefield fluorescence microscopy. Lastly, powerful mathematical tools are available for pattern evaluation, clustering, or co-localization studies. The most important of these tools will be highlighted in the following chapter 1 to 2.7.

2 METHODS & METHOD DEVELOPMENT FOR SMLM DATASETS

Single-molecule localization microscopy (SMLM) converts a diffraction-limited movie of fluorescence microscopic images into a list of analyzable coordinates. Since this process involves various steps, each with its own parameters that need to be optimized, it is important to ensure that the final reconstructed super-resolution image displays the actual underlying sample. Therefore it is necessary to take a closer look at image quality and resolution. When evaluating the resolution of SMLM data, the workflow can be divided into three categories. A) The hardware. This includes the microscope and the attached illumination scheme (laser excitation, objective, stage, filter, and camera). B) A suitable sample for testing image resolution. This includes an appropriate sample design and choice of light emitter (like fluorescent proteins or fluorescent dyes). C) The data analysis that includes the whole image reconstruction routine and optional post-processing steps.

Hardware. The most essential part for SMLM is naturally the microscope itself. In order to resolve single-molecules and to localize them with a precision of a few nanometers, it is of particular importance that the microscope setup is properly designed and aligned to obtain the highest resolution possible. The setup used throughout this work is a commercial Nikon N-STORM microscope. This microscope bundle includes specific components chosen to optimize SMLM performance. Most notably are:

- A high-power laser excitation (i.e. a 647 nm laser with 125 mW output power at the tip of the fiber) that allows a high photon flux and results in increased photon statistics and rapid imaging.
- A 100x oil objective with a numerical aperture of 1.49 as well as a highly sensitive EMCCD camera (Andor Ixon Ultra 897) which ensures that as few photons as possible are lost during imaging and maximizes the signal noise ratio.

A detailed listing of the hardware components is available in chapter 8.1. Appropriate literature was followed to build up and optimize the microscope for SMLM and to choose the most suitable components [45], [46]. Therefore, a detailed analysis of this aspect is not within the scope of this doctoral thesis.

Sample. In order to work with a controllable and reproducible sample, a “test sample” has to be introduced. The general design of test samples to evaluate the resolution of the microscope is very flexible and multiple approaches are available. Fluorescent beads and quantum dots can be used very effectively to quantify stage drift [42]. Crossing labeled microtubules are often used as a biological sample to evaluate resolution [47], however the exact shape of these structures is arbitrary and not reproducible. On the contrary, synthetic nanostructures based on DNA origami can be modified with the fluorophore of choice at nanometer specific positions [48]. Therefore, DNA origami was used to evaluate the microscope resolution as presented in chapter 3 of this thesis.

Analysis. The analysis of SMLM data comprises various steps. First of all, errors in the attained SMLM images like sample drift or excessive noise have to be excluded and successful spot-finding and image reconstruction have to be ensured. These requirements will be addressed in chapter 2.1, where the super-resolution reconstruction workflow that has been used in this doctoral work will be introduced. In a second step, the final image is evaluated concerning resolution and image quality (see 2.2). Finally, additional aspects can be analyzed in order to answer specific questions of interest, like segmentation (1), clustering (2.4) and co-localization (2.5). Chapter 2 will conclude with the introduction of two novel analysis tools: SiMoNa – a graphical user interface (GUI) for exploring and analyzing SMLM datasets in MATLAB (2.72.6); and PairRice – an accurate pairwise distance model and fitting procedure (2.6).

2.1 Super-resolution reconstruction

In order to reconstruct a super-resolution image from an SMLM movie, various open-source software packages are available. In this work, ThunderSTORM [49] has been used since it performed superior in a publication comparing and evaluation super-resolution software [47]. Additionally it is built into the widely used ImageJ platform. ThunderSTORM features multiple methods and tools to reconstruct the super-resolution image. The underlying workflow and its key parameters will be introduced now:

Spot finding and fitting. As a first step, image filtering is used to enhance the features to be detected (i.e. the fluorescent spots). The first level of an “à trous” undecimated wavelet transform with a B-spline order of 3 and scale of 2 contains high frequency components of the image. Its standard deviation will be used for single-molecule detection [50]. The spot finding is performed with a local threshold in an 8-connected neighborhood based on aforementioned standard deviation. For the sub-pixel localization, the expected point-spread function of a single emitter is estimated by a symmetric 2D Gaussian in its integrated form to account for the discrete pixel size. The final fitting uses maximum-likelihood methods to minimize the sum of the squared residual. ThunderSTORM then outputs a list of localizations of which the following parameter are the most important ones:

- Position (x, y) in nm
- Recorded frame F
- Photon intensity I (as the integral over the whole fitted 2D Gaussian)
- The width (standard deviation) σ of the fitted 2D Gaussian

Drift correction. Traditionally, software-assisted drift correction in fluorescence microscopy is performed with the help of so called “fiducials” after imaging. These bright fluorescent markers like quantum dots, gold nanoparticles, or fluorescent beads are immobilized at a low density and anchored to the surface such that their positions can be tracked with high precision. The drift is corrected by subtracting any deviations from its initial position over the course of the recording. However, the very bright fluorescence emission of these markers may outshine and thus interfere with detection of the surrounding sample. Furthermore, an additional surface preparation step to immobilize the fiducials is necessary.

Therefore, an alternative approach has been developed and modified for SMLM that relies on cross-correlation of the recorded emitter localizations [51]. It splits all localizations into multiple sub-populations based on detection frame intervals. A super-resolution image is reconstructed for each sub-population and cross-correlations between the first and the following super-resolution images are calculated. Every feature (i.e. localization) in the super-resolution image at time zero at any position (x, y) will correlate to the same feature in the image at time t at position $(x + D_x, y + D_y)$, where D_x and D_y correspond to the drift in the x and y direction, respectively. Therefore, the cross-correlation images show a central peak at the position of the drift vector (D_x, D_y), as shown in Figure 2.1A. The peak positions can then be plotted and interpolated for all frames to yield a continuous drift correction curve (Figure 2.1B).

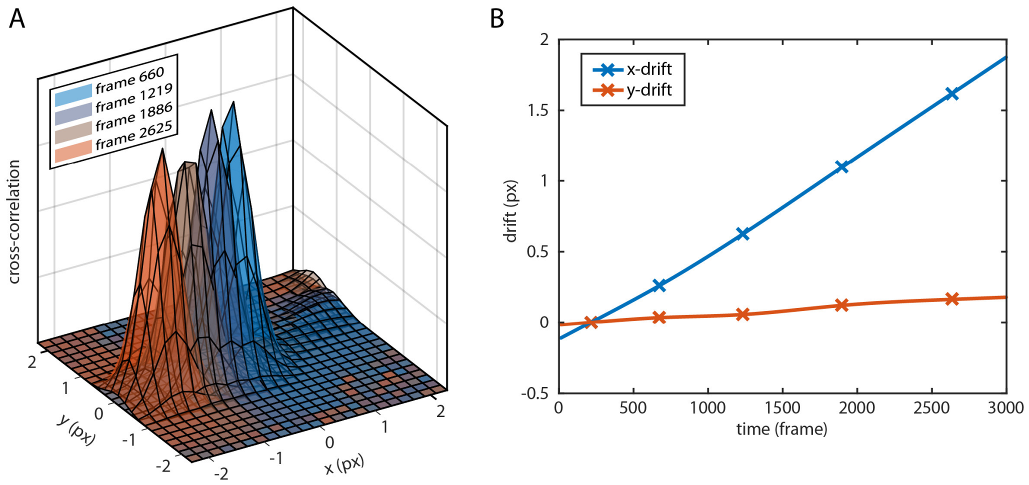


Figure 2.1 Drift correction procedure based on cross-correlations A) Cross-correlation intensities for different frames. The peak position represents the drift vector. B) The resulting drift correction curve is generated by interpolating between the discrete cross-correlation values.

Image reconstruction. Multiple methods have been established to convert a list of coordinates into a super-resolution image with a user-specified magnification [52]. A straight-forward approach is to create a histogram that bins all localizations into a reasonably small bin size to create a super-resolved image. This will not include the localization precision [53] per molecule as all localizations are treated equally as binary elements. Therefore, the most commonly used method is to represent each localization not as a single dot but as a 2D Gaussian with the standard deviation equal to the localization precision. All Gaussians are added up and presented in a single image with a user-specified resolution.

After the successful image reconstruction it is now possible to evaluate the final image with regards to resolution and quality, i.e. to ensure that the image is free of artifacts.

2.2 Tools for SMLM resolution estimates

This section describes the tools to evaluate SMLM datasets in a global fashion, i.e. using the whole dataset of the raw localizations. Therefore, these tools can be applied to any dataset, irrespective of the sample, to gain initial insights into its resolution properties. In comparison, local resolution estimates are applied in specific areas and properties of a certain sample and are thus highly specialized. For the specific case of DNA origami, some local resolution methods tailored to this specific experiment design will be introduced later on in chapter 3.3.

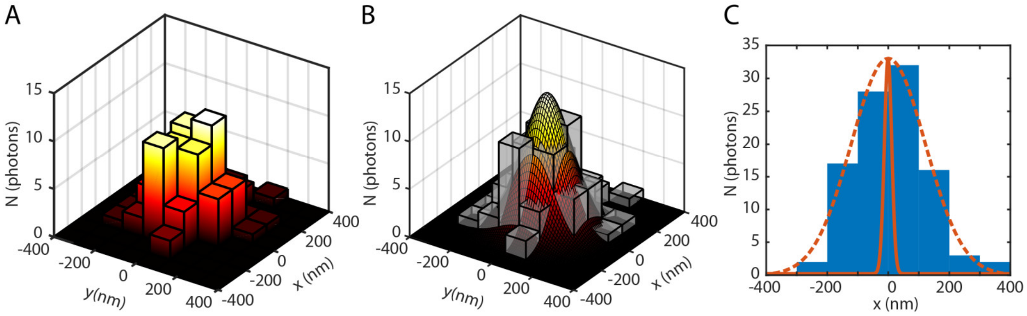


Figure 2.2 Fitting single emitters with a 2D Gaussian A) Pixelated intensity distribution of a single emitter as imaged by the EMCCD camera. B) The intensity distribution can be approximated by a 2D Gaussian. C) The fitting (dashed red curve) yields a precise centroid measurement (solid red curve).

Localization precision. The immense resolution gain of SMLM stems from the fact that a single emitter – estimated as a 2D Gaussian profile (Figure 2.2A & B) – can be fitted very precisely and thus its can be peak determined with high accuracy (Figure 2.2C). The most straight-forward approach for a resolution estimate would hence be:

$$\langle (\Delta x)^2 \rangle = \frac{\sigma^2}{N} \tag{3}$$

where σ^2 is the width (variance) of the fitted Gaussian and N is the number of detected photons within the peak. Thompson *et al.* [53] refined this formula to account for the finite pixel size a and the background noise that is always present as a dark count b for the detector:

$$\langle (\Delta x)^2 \rangle = \frac{\sigma^2}{N} + \frac{a^2/12}{N} + \frac{8\pi\sigma^4 b^2}{a^2 N^2} \tag{4}$$

For the Andor Ixon Ultra 897 that is part of the Nikon N-STORM system the pixel size a for a 100x objective is 106.7 nm and the dark count b is 89 electrons (which is reduced below one electron in electron multiplication mode). Both adjustment terms reduce the localization precision Δx. Therefore, a camera with a small pixel size and a low dark count is to be preferred.

Nyquist sampling theorem for density evaluation. It is important to emphasize that when attempting to image and to reconstruct a certain structure with high resolution, it is not sufficient to achieve a high localization precision per emitter. Since we are working with a pointillist approach of image reconstruction, the localization density is a crucial parameter. The Nyquist sampling theorem [54] states that a signal can be completely recovered from its samples if the sampling frequency is at least twice the highest frequency in the original signal. This means that two data points are necessary per resolution unit. This resolution unit is normally limited by the pixel size of the imaging setup. In the case of SMLM it is equal to the localization precision. In order to satisfy the Nyquist theorem, the following requirement has to be met:

$$molecular\ density \geq (2/\Delta x)^d \tag{5}$$

with d being the dimension. For the purpose of illustrating this criterion, we assume that the average localization precision is 20 nm (a reasonable value for a PALM experiment) and that the structure of interest lies in a 1 x 1 µm field of view in 2 dimensions. In this case, we need at least 10’000 localizations in said area to satisfy the Nyquist criterion. Else, the image quality is degraded and the image may show gaps which might be misinterpreted as actual gaps in the imaged structure yet are indeed just artifacts due to the sparse localization count.

The following simulation illustrates this effect: A binary mask is created to distribute N points randomly in a pattern of choice (Figure 2.3A - left). Two different localization precisions Δx for all points have been chosen (5 and 20 nm) as well as different number of points N to illustrate different scenarios. A low localization precision will not allow fine features to be resolved, even at high localization densities (Figure 2.3A - top row, Figure 2.3B – blue curves). A low localization density on the other hand will result in sparse patterns with arbitrary shapes that do not represent the underlying real distribution (left column of Figure 2.3A and dark dotted curves in Figure 2.3B). Only the combination of a sufficiently high localization density and precision will yield an accurate representation of the underlying structure (bottom right of Figure 2.3A).

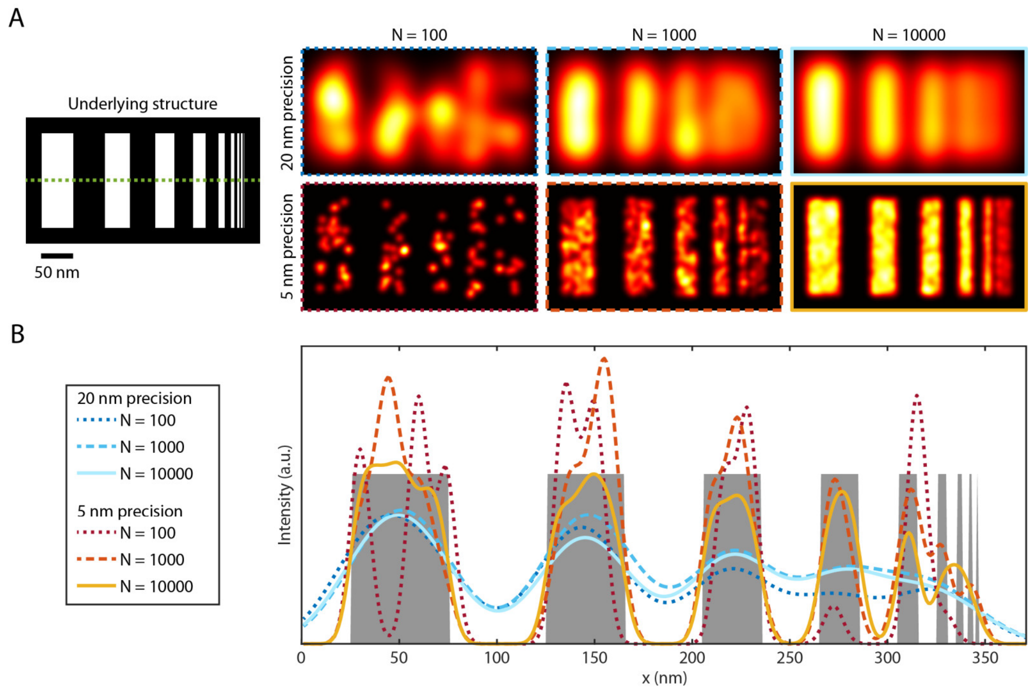


Figure 2.3 The effect of localization density and precision A) The underlying structure acts as a binary mask where localizations will be randomly distributed. The green line represents the region for the line scan shown in B). Reconstructed super-resolution images for different localization precisions and number of localizations N are shown on the right. B) Line scans for all six scenarios are displayed with the underlying structure represented as gray boxes.

While the particular number of localizations needed is highly dependent on the structure and biological question addressed, it is worth noting that the higher the localization precision, the higher the localization count has to be. Obviously, imaging in 3 dimensions requires a higher localization count as well.

Fourier ring correlation for a global resolution estimate. The Nyquist theorem builds a value for an average density. It is thus not able to incorporate the actual localization pattern with a heterogeneous density. Therefore it is necessary to correlate each localization with its unique local environment. For this purpose, an algorithm based on Fourier ring correlation (FRC) is widely used [55], [56]. Two independent reconstructions of the same object are calculated and their similarity is evaluated to determine a resolution threshold where both reconstructions are still consistent. Up to this resolution (which equals a certain spatial frequency) the object can be considered as resolved. Mathematically, the FRC calculates the degree of correlation between two images for different spatial frequencies. For SMLM in particular, this method is applied as follows:

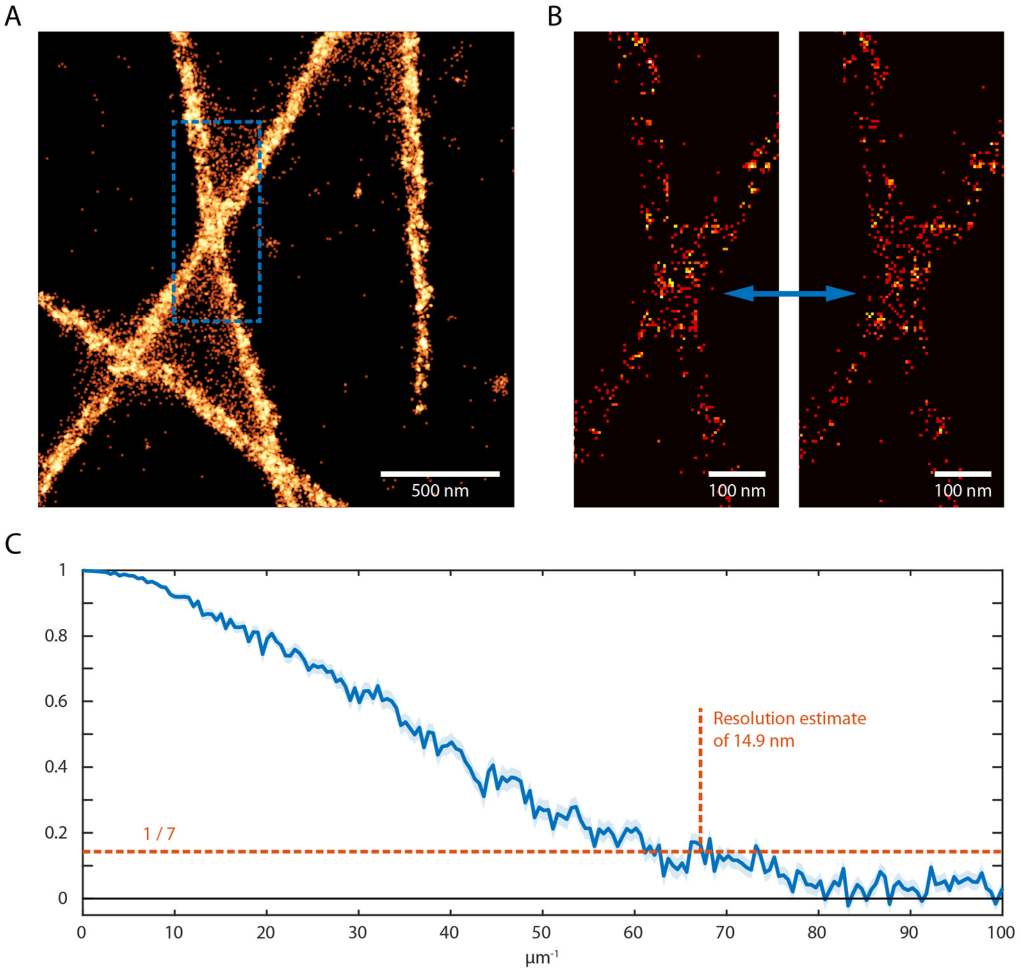


Figure 2.4 Fourier-ring correlation workflow A) Super-resolution scatter plot of filamentous structures. B) Two super-resolution images used for the cross-correlation, consisting of 50% of all localizations each, showing the region highlighted in A) reconstructed using a bin size of 5 nm. C) Fourier-ring correlation yielding a resolution estimate of 14.9 nm for a threshold of 1/7.

1. The localizations (Figure 2.4A) are split into two equal (but randomly selected) populations and are binned into two super-resolution images (Figure 2.4B). The bin size has to be sufficiently small as it sets the lower limit for the resolution for the resolution estimate.
2. The two images are then Fourier transformed and multiplied.
3. The average correlation of multiple concentric rings (corresponding to multiple frequencies) is calculated and will give rise to the FRC curve for the SMLM dataset (Figure 2.4C).

4. Via a threshold in the FRC curve, a resolution can be defined. Nieuwenhuizen *et al.* found the threshold of 1/7 is most appropriate for SMLM images [56].

2.3 Voronoi tessellation for noise-removal and cluster estimation

In general, clustering means grouping a set of objects into subsets such that each object of the same cluster has similar properties. In fluorescence microscopy, where the molecules of interest are distributed in two or three dimensions, the property that defines clustering will be the spatial dimension. For SMLM and particularly in this work, it is important to determine whether or not the x and y localization coordinates of tagged proteins are distributed in clustered patterns.

When working with potentially clustered SMLM data the most challenging first step in data analysis is to determine whether the data is actually clustered while using an approach that does not need custom parameters or thresholds that might influence the result of the analysis. Most of the tools described in the last chapter are very helpful to quantify the dataset if it is assured that the studied structure is indeed clustered and even more so if the underlying structure is known (e.g. in the case of the DNA origami, as shown in chapter 3). For an unexplored biological sample, e.g. the cingulins studied in in this work (chapter 6), this is hard to predict.

This chapter focuses on a recently developed method based on Voronoi tessellation [57], [58]. Tessellation in general describes the tiling of a plane (i.e. localizations in 2D) into geometrically defined shapes. Voronoi tessellation does not need any input parameters and serves as an initial step to segment and quantify the dataset and thus helps in identifying outliers and noise in the data, clustered areas, and the general shape of the biological structure.

The starting point is a super-resolution image as shown in Figure 2.5A. Out of all the localizations of interest, a Voronoi diagram is formed (gray lines in Figure 2.5B). Voronoi diagrams are a tessellation technique that partitions a plane into N polygonal regions (Voronoi cells) based on a point cloud consisting of N points. Each Voronoi cell surrounds a single point $P(x, y)$ and forms an area around it where P is the nearest point from all points inside the cell. The area of each polygon can be easily calculated and gives a first insight in the local point density around each point (color map in Figure 2.5C). By looking at the polygon area distribution and applying an area threshold it is already possible to differentiate high *versus* low density areas

(Figure 2.5E). Of course, this threshold is arbitrary so far. In order to determine an objective threshold it is necessary to compare the experimental data with a random dataset. This dataset consists of the same amount of points distributed randomly over the same area as in the experiment (Figure 2.5D). By applying the same Voronoi diagram analysis to said dataset, it is possible to compare the polygon area distribution (Figure 2.5E) since they have the same data size. For the highly clustered example data used in Figure 2.5, two intersections in the histogram form two thresholds. These two thresholds segment the dataset into three species as shown in Figure 2.5F. Below the first intersection of 750 nm^2, the experimental data points are more likely to be clustered than random points. Between the first and second intersection at 9000 nm^2 the points do not belong to a cluster and are in a state of random distribution. Above the second intersection the points can be interpreted as noise. After this segmentation, one can remove the points flagged as noise or further analyze the identified and cleaned up clusters, for example with density-based clustering techniques like DBSCAN.

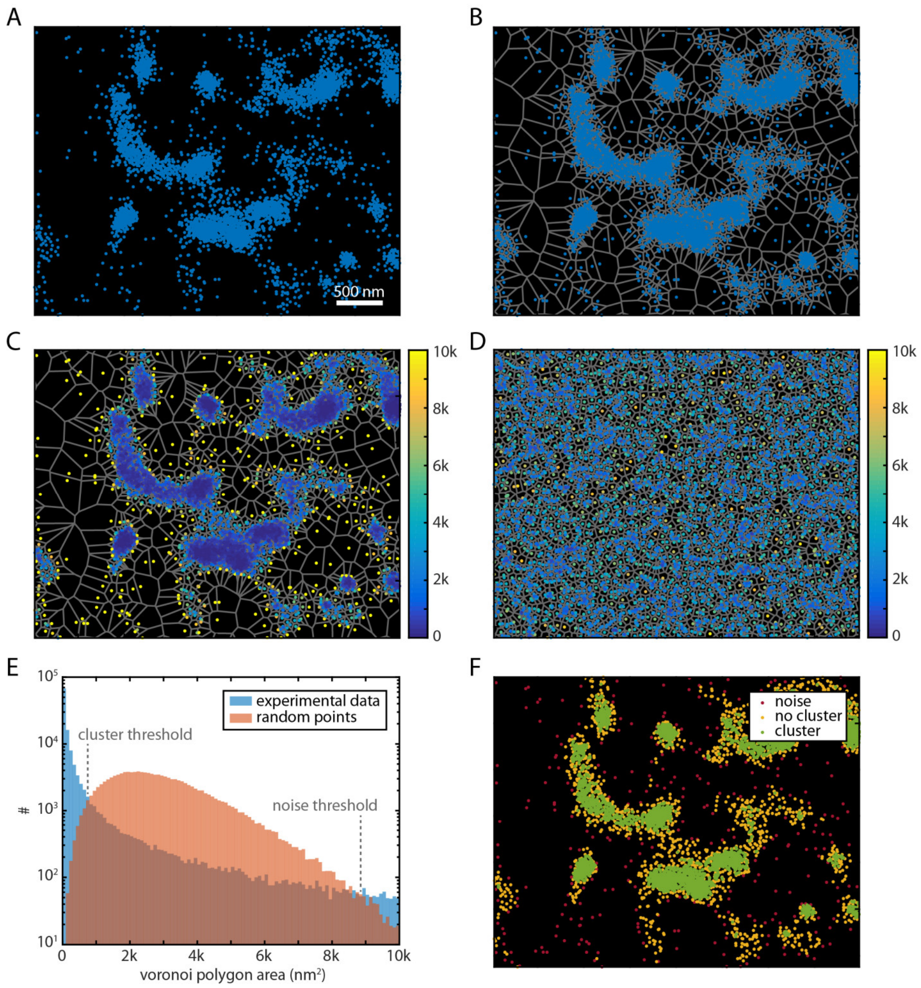


Figure 2.5 Overview of the Voronoi workflow A) Example of clustered raw localizations. B) The calculation of the Voronoi diagram connects each localization to a polygon (gray). C) Localizations are colored by their surrounding polygon area in square nanometers which clearly shows dense areas in dark blue. D) The same number of points are shown in a randomly distributed set that acts as a reference. E) The histogram of the Voronoi polygon area of both datasets reveals thresholds for clustered points (below the first intersection) and noise (above the second intersection). F) By applying these thresholds to the dataset one can distinguish between clustered data (green), non-clustered data (orange) and background noise (red).

2.4 Tools for SMLM cluster analysis

In order to evaluate clustered SMLM data, multiple tools are available that can help quantify the data. Here, the most common ones will be described very shortly with their most important field of application. The application of the algorithms that are relevant to this work will follow in the result chapters (starting with DNA origami in chapter 3). It is worth mentioning that all the algorithms listed here work on point clouds (i.e.: raw single-molecule localizations) which is a big advantage compared to the “classical” clustering techniques that were established in the field of fluorescence microscopy which rely on an intensity-based pixelated finite image.

Table 2 lists the algorithms that will be described in detail in the following paragraphs and highlights whether an algorithm requires the number of clusters k in a dataset as an input parameter or if it yields it as an output parameter. The first two algorithms, Ripley’s K function and pairwise distance distribution, will not yield specific information on the number of clusters in the dataset or on the assignment of a certain molecule to a specific cluster. Nevertheless, they are of importance as Ripley’s K function can be applied to a whole dataset to estimate if it is clustered at all while the pairwise distance distribution on the other hand allows precise distance calculations when it is known how many clusters are present.

Table 2 Overview of clustering algorithms Important clustering algorithms used in this work are listed. k refers to the number of clusters.

	k as input necessary	*k as output available*	*correlates molecules with cluster*
Ripley’s K function	No	No	No
Pairwise distance dist.	No	No	No
DBSCAN	No	Yes	Binary
k-means clustering	Yes	-	Binary
Gaussian mixture dist.	Yes	-	Probability

Therefore, Ripley’s K function will be used throughout this work for a first estimation if and how a dataset is clustered, both for synthetic nanostructures like DNA origami (chapter 3) as well as for biological samples of unknown geometry (chapter 6). Pairwise distance distribution, k-means clustering and Gaussian mixture distribution are compared extensively on DNA origami nanostructures (chapter 3) to extract inter-cluster distances as the number of expected clusters is known for this sample. DBSCAN is able to extract the number of clusters k autonomously. Thus it is applied to the images of cingulins in chapter 6 to count and quantify multifacetedly shaped clusters.

Ripley’s K function and its derivatives. A global method to evaluate if a certain point distribution is clustered is given by the functions developed by B. D. Ripley

[59]. $K(r)$ describes the amount of neighboring points within a radius r around each other, normalized by the density:

$$K(r) = \frac{A}{N}\sum_{i\neq j}\frac{I(d_{ij}<r)}{N} \qquad (6)$$

A is the region of interest, N is the total number of points, d_{ij} is the distance between point i and j, r is the search radius and I is the indicator function. For a single point and search radius r, the algorithm is visualized in Figure 2.6. By comparing randomly distributed points in panel A) with clustered points in panel B), the resulting change in the K function for certain radii r becomes apparent.

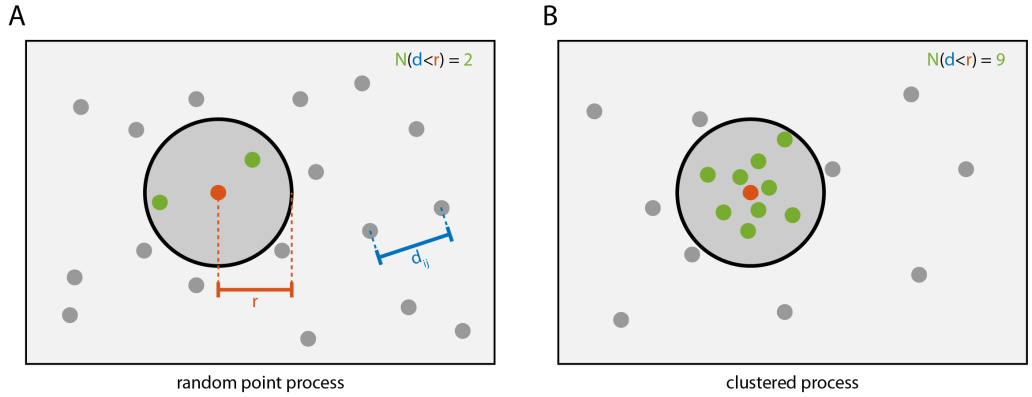


Figure 2.6 Ripley's K function used to analyze spatial point patterns A) For a single point of a random point process, the search radius r is shown. The points within this radius are displayed in green. B) For the same number of total points in a clustered pattern, the number of points within search radius r increases.

For randomly distributed points (as defined by a homogeneous Poisson point process), the K-function is described by:

$$K_P(r) = \pi r^2 \qquad (7)$$

Any deviation from this random process towards higher or lower $K(r)$ indicates a higher (clustering) or lower (dispersion) density at certain radii r, respectively. The K-function can further be normalized to simplify the random process:

$$L(r) = \sqrt{K(r)/\pi} \qquad (8)$$

$$H(r) = L(r) - r \qquad (9)$$

The H-function can be used efficiently in estimating clustering, as a random process is expected to stay at zero for all radii. Therefore any peaks above or below zero at radius r display clustering or dispersion at that length scale, respectively [60].

Pairwise distance distribution. A valuable tool to extract distances between two clusters is the "pairwise distance distribution" between all points that works even when the two clusters are strongly overlapping. The expected distribution features two populations: one describes distances within a cluster (see Figure 2.7, red lines). Its peak position reflects the width of the individual clusters. The second population with its peak at a bigger distance describes the distance between the two clusters (Figure 2.7, green lines).

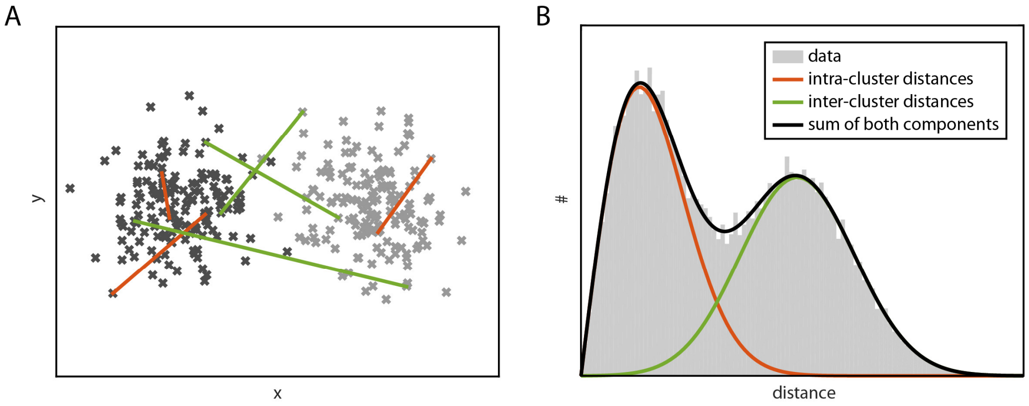


Figure 2.7 Pairwise distance fitting scheme A) Simulation of localizations of two clusters in close vicinity. Distances within a cluster are marked in red, distances between the clusters in green. B) Histogram of all distances displays two populations of intra- and inter-cluster distances.

An approximation of this distribution for fitting purposes has been developed by Tinnefeld *et al.* [48] which assumes two underlying distributions: a 2D Maxwell-Boltzmann $v(x)$ and a Gaussian $g(x)$ for a certain distance x:

$$f(x) = v(x) + g(x) = ax \cdot \exp(-bx^2) + h \cdot \exp\left(-\frac{(x-d)^2}{2\sigma^2}\right) \qquad (10)$$

Here, b and σ depend on the width of the clusters. a and h are scaling parameters and d is the cluster distance. This method is able to accurately calculate the distance between two point emitters (that create a Gaussian shaped distribution per cluster). However, its limitations are:

- The second term $g(x)$ is only an approximation and will not hold true for overlapping clusters with small distances.
- Most of the fitting parameters do not have a direct physical equivalent like localization precision or number of localizations per cluster.
- The fitting parameters b and σ as well as a and h are connected to each other, therefore the number of fitting parameters can be reduced.

In order to overcome these limitations, a new approach has been developed in this doctoral work which is described in chapter 2.6.

DBSCAN. Density-based spatial clustering of applications with noise (DBSCAN) is probably the most used density-based clustering algorithm [61]. It is based on connecting points within a certain distance ε and grouping them together. Points that cannot be connected to a minim amount of neighbors $minPts$ within the distance ε will be considered as outliers / noise. Hence, DBSCAN has two input parameters (ε and $minPts$) that define thresholds that have to be chosen manually and will ultimately assign each point to a certain cluster or to background noise.

k-means clustering. k-means clustering is a centroid-based clustering method that dates back until 1957 [62]. Its main drawback is its need of the number of clusters as an input parameter. The method tries to find the k cluster centers by assigning each point to the closest cluster center and minimizing the squared distances from all points to these cluster centers. Since this is an iterative process that relies on the starting conditions, the algorithm may only find a local optimum and multiple runs might yield different results. Moreover – since each point is assigned to the nearest center – clusters tend to be of similar size.

Gaussian mixture distribution. The Gaussian mixture distribution belongs to the class of distribution-based (or model-based) clustering algorithms. Its initial assumption is that each cluster can be defined by a certain distribution (in the case of SMLM data a 2D Gaussian distribution) and therefore all data points can be fitted to a certain underlying global probability distribution. This fit will be achieved as many other fits by an iterative optimization procedure (most prominently the expectation-maximization algorithm [63]). As more fitting parameter (like more clusters, more flexible distributions) will always results in a better fit, it is important to constrain the model and to fix the number of clusters to be fitted. Often, distribution-based clustering will be combined with other clustering methods (like k-means++ [64]) that give reasonable initial values for the number of clusters and the localizations of the cluster centers. An advantage of this algorithm is its flexibility in the choice of the underlying distribution and its simple adjustment to fit the experimental conditions. Additionally, this method yields a cluster probability for each individual point instead of a binary correlation of a point to a certain cluster, On the other hand it suffers from the same drawbacks as k-means clustering with the results depending on the starting conditions.

With these clustering tools at hand it is now possible to evaluate SMLM datasets from initial clustering guesses with the help of Ripley's K function, to precise quantitative analysis with DBSCAN.

2.5 Coordinate-based co-localization

Finding correlations between multiple species of labeled biomolecules is an important aspect of fluorescence microscopy, especially when the interaction of different molecules / proteins is studied, e.g. the co-localization of different cingulins in diatoms (chapter 6). In traditional widefield microscopy, co-localization studies could be performed qualitatively by simply overlaying the images in multiple colors or quantitatively by comparing the intensity values of the same pixels for the different colors [65]. With the help of SMLM, correlations between different species are possible with higher precision. The mentioned qualitative process of overlaying the multiple super-resolution images is still possible. However, the quantitative correlation process has to be adjusted to be compatible with the single-molecule localization data. For this purpose, the so called coordinate-based co-localization (CBC) has been developed [66], [67]. A co-localization value C is created for each localization from each of the two channels. The calculation follows a similar scheme like Ripley's K function. Its workflow will be described briefly in the following:

1. For a certain localization i in channel A, the number of localizations in the same channel within multiple distances r around A are calculated (up to a maximum distance R_{max}).
2. Additionally, for this localization i in channel A, the number of localizations in channel B within distances r around A is calculated as well.
3. These distributions depending on r are corrected for the area and normalized by the average localization density.
4. Finally Spearman's rank correlation coefficient between those two distributions is calculated.
5. This value is then normalized to fall between -1 (anti-correlation) and $+1$ (correlation), which yields the final co-localization value C.
6. This value C is calculated for each localization i in each channel.

For illustration purposes an exemplary simulation has been carried out for two simply-shaped overlapping populations (Figure 2.8A). In panel B and C, the co-localization value C for each channel is plotted. The region of overlap is recognizable by high C-values.

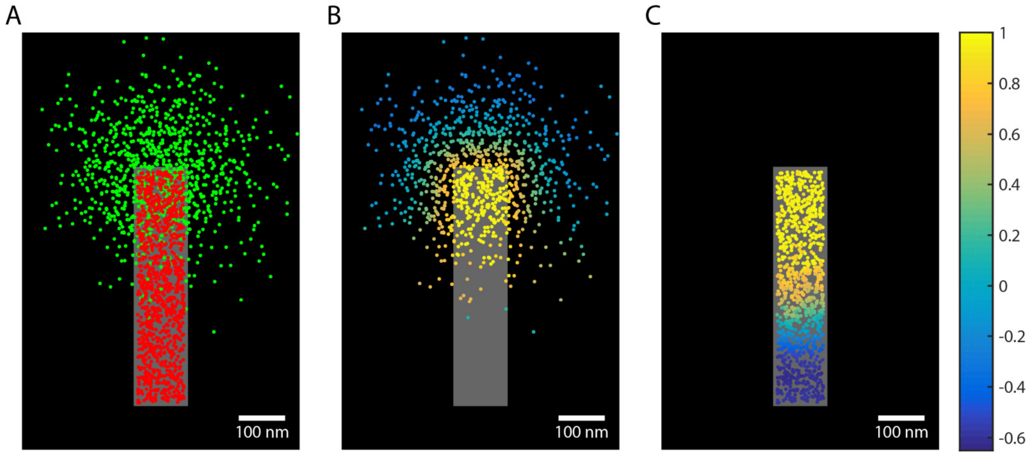


Figure 2.8 Coordinate-based co-localization for simulated data A) Two point clouds of 1000 points each have been randomly distributed as a normal distribution (green) and in a rectangular shape (red). B) & C) Co-localization values C as calculated by the CBC algorithm for the green and red channel, respectively. Both plots clearly show regions of high co-localization in yellow.

The only input parameters needed are the maximum search radius R_{max} and the binning radius Δr that defines the step size for the radii r that are probed. The authors of the original publication recommend a maximum search radius in the range of the expected structures that are probed and a binning radius approximating the resolution of the PALM / STORM image.

2.6 PairRice – A novel algorithm to extract distances between cluster pairs

An approximate pairwise distance formula to measure distances between a pair of clusters has already been introduced in chapter 2.4. However, its major limitation – which is the fitting of just two populations – may be severe when looking at arbitrarily shaped structures in biological SMLM images. For studies where the underlying structure is known – as it is the case with DNA origami in chapter 3 – this is not a concern. The usability of the formula presented in 2.4 is also limited as it will not hold true for overlapping clusters and contains superfluous fitting parameters with most of them not having a physical equivalent. These drawbacks will be overcome with the PairRice algorithm which I developed within the scope of this work.

With PairRice I developed an accurate and stable method to determine the distance between two single-point emitters in a SMLM dataset. The original formula is given by:

$$f(x) = v(x) + g(x) = ax \cdot \exp(-bx^2) + h \cdot \exp\left(-\frac{(x-d)^2}{2\sigma^2}\right) \qquad (11)$$

Here, the inter-cluster peak $g(x)$ has been simplified to a Gaussian distribution. This simplification is not exact for small distances between the clusters, as the probability has to be zero at the distance $x = 0$, which is not the case for a Gaussian. Therefore, an exact solution had to be developed. An additional motivation was to replace the arbitrary fitting parameters with physical meaningful representations. Starting point is the Rice distribution R that describes the probability for a circular bivariate normal random variable (2D Gaussian, with a width σ) with a non-zero mean d:

$$R(x|d,\sigma) = \frac{x}{\sigma^2} \exp\left(\frac{-(x^2+d^2)}{2\sigma^2}\right) I_0\left(\frac{xd}{\sigma^2}\right) \qquad (12)$$

Here, I_0 is the modified Bessel function of the first kind with order zero. Both peaks in the expected distribution can be described by such a distribution:

$$R(x|d,\sigma) = R_1(x|d,\sigma) + R_2(x|d,\sigma) \qquad (13)$$

R_1 describes the intra-cluster peak (red in Figure 2.7), therefore $d = 0$ and the distribution becomes a Rayleigh distribution:

$$R_1(x|0,\sigma) = \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) \qquad (14)$$

As a next step, a proper normalization factor has to be introduced. The area under the curve R scales with the number of pairwise distances N. n localizations always result in:

$$N = \sum_{i=1}^{N-1}(N-i) = n\,(n-1)/2 \qquad (15)$$

pairwise distances. Unfortunately, the two cluster are not equally populated, resulting in different ratios between the two peaks R_1 and R_2. Therefore a fitting factor k has to be introduced that represents this ratio and is 1 for equal cluster size, where there is the same amount of intra- and inter-cluster distances:

$$R(x|d,\sigma) = \frac{N}{2} k \cdot R_1(x|0,\sigma) + \frac{N}{2}(2-k) \cdot R_2(x|d,\sigma) \qquad (16)$$

Finally, the width of both distributions σ has to be calculated. If a single cluster is normally distributed and its width (standard deviation) is given by the average localization precision $\langle \Delta x \rangle$, then the difference between two points is also normally distributed and their variances add up:

$$\sigma^2 = 2\langle \Delta x \rangle^2 \qquad (17)$$

Since the average localization precision of all events is just an estimate for the true width of the two clusters and does not include heterogeneities, this parameter is also a free fitting parameter. This results in the final fitting formula:

$$R(x|N,k,d,\Delta x) = \frac{N}{2}kR_1\left(x|0,\sqrt{2}\Delta x\right) + \frac{N}{2}(2-k)R_2\left(x|d,\sqrt{2}\Delta x\right) \quad (18)$$

For fitting these populations, reasonable starting parameters are:

- $k_0 = 1$.This would correspond to two equally populated clusters.
- $\Delta x_0 = \langle \Delta x \rangle$. A good first guess for the width of the fit is the average localization precision.
- $\mathrm{d}_0 = \langle x \rangle$. For the actual unknown cluster distance, the average pairwise distance is a reasonable starting point for fitting.

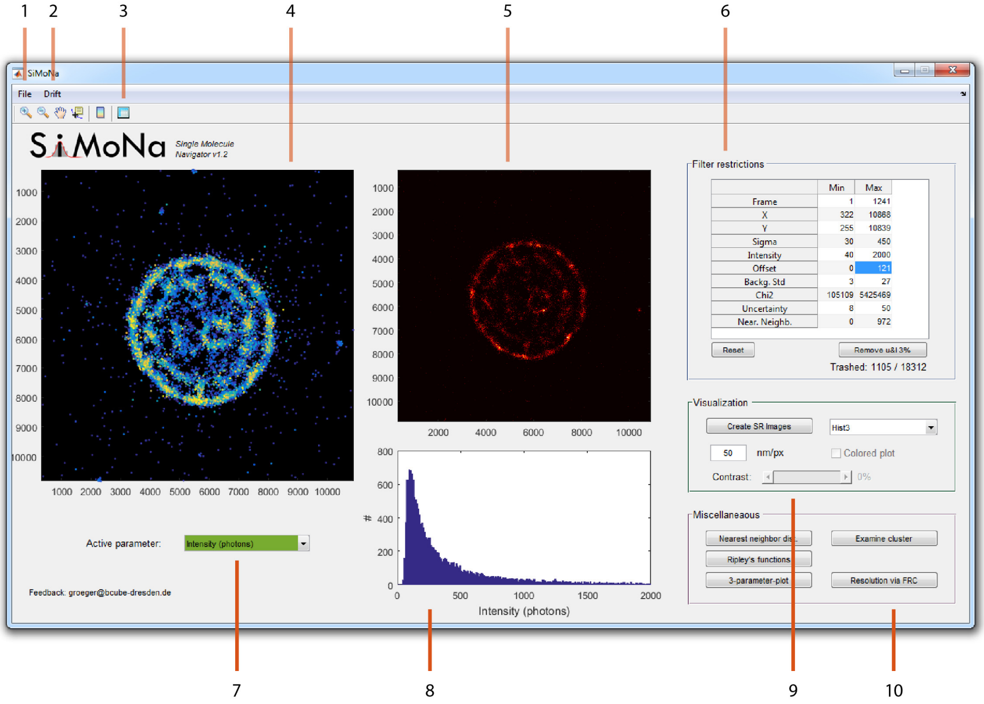
The resulting fit formula can be applied to the clustered data via standard fitting methods like a nonlinear least-squares solver.

In summary, the novel PairRice fit reduced the free parameter from 5 to 3, thus making the fitting procedure more stable as compared to the original formula. The very good performance of the fitting method was confirmed by applying it for resolution estimates of the SMLM setup (presented in chapter 3).

It has to be noted that PairRice is not an approximation but in fact the accurate description of the pairwise distance distribution for a pair of Gaussian shaped point clouds in general. Thus it can not only be applied to SMLM but to all clustering and distance measurement tasks that fall into this category.

2.7 SiMoNa – A new GUI for exploring SMLM datasets

A big advantage of SMLM that is often underestimated is the continuous access to the raw data / localizations that formed the reconstructed super-resolution image. Therefore it is possible to use the localization parameters for further evaluation (by visualizing the parameter distributions) and filtering (by applying parameter thresholds). In the ThunderSTORM software package, data filtering as well super-resolution image reconstruction can be performed in different ways. However, an easy-to-use, elaborate and flexible graphical user interface (GUI) to explore these datasets has not been available to date. Therefore, I developed a MATLAB GUI that allows loading of localization files to further display, filter and analyze them (e.g., for resolution estimates, clustering, and more). This chapter introduces SiMoNa: the **Si**ngle **Mo**lecule **Na**vigator. The software is available upon request for academic institutions. This section will introduce the standard workflow of data analysis and highlight novel features of SiMoNa to explore SMLM datasets.



	Min	Max
Frame	1	1241
X	322	10868
Y	255	10839
Sigma	30	450
Intensity	40	2000
Offset	0	121
Backg. Std	3	27
Chi2	105109	5425469
Uncertainty	8	50
Near. Neighb.	0	972

Trashed: 1105 / 18312

Hist3

50 nm/px

Contrast: 0%

Active parameter: Intensity (photons)

Feedback: groeger@bcube-dresden.de

Figure 2.9 SiMoNa main window The main GUI window with the most important features highlighted. 1) File import and export options. 2) Options to remove linear drift or load a drift file from an external source. 3) Display of all localization with all their parameters in a table. 4) Scatter plot with the localizations colored by the active parameter. 5) Preview of the resulting super-resolution image. 6) This table displays extrema of all parameters and allows filtering by applying parameter thresholds. 7) Drop-down menu to choose the active parameter to display. 8) Histogram of the active parameter. 9) Multiple options to form a super-resolution image, including coloring by the active parameter. 10) Additional display and analysis options. See the main text for details.

Data import. SiMoNa supports the import of ThunderSTORM localization files. Additionally, already previously imported and analyzed data can be imported as MATLAB data structures for further analysis. Further import scripts are simple to generate so that the use of SMLM datasets from other software can be easily introduced if required.

Data exploration. A unique feature of SiMoNa is the ability to display the dataset in a scatter plot, colored by a localization parameter of choice – which is in the following called “active parameter – like frame, brightness or goodness of the Gaussian fit (Figure 2.9, left plot). Simultaneously, a histogram of the active parameter is shown and multiple localization parameter can be plotted against each other as well. This is a huge advantage over many other super-resolution image reconstruction software programs (like ThunderSTORM, RapidSTORM [68], or QuickPALM [69]) that mainly offer monochromatic super-resolution reconstruction routines. Therefore, the spatial distributions of localization parameters remain hidden. For illustration purposes, Figure 2.10 displays a scenario where "faulty" localizations (e.g. by setting wrong ThunderSTORM spot detection parameters) can be identified and filtered out. Here, the super-resolution scatter plot, colored by the localization precision, displays inhomogeinities (Figure 2.10A). The histogram of the localization precision reveals the presence of multiple populations (Figure 2.10B). By plotting the sigma and the offset of the fitted Gaussian against each other (Figure 2.10C), it becomes apparent that the two outlier-populations are identifiable by a sigma below 10 nm and an offset below 20 photons (which coincides with a very high sigma above 500). In this example, the low sigma could be explained by a fitting of single noisy or hot pixels of the camera (which therefore results in a very sharp Gaussian with a low sigma value). The low offset and high sigma, on the other hand, could stem from a sample whose fluorophore density is too high to accurately fit single emitters. Therefore, ThunderSTORM may have introduced faulty localizations by fitting multiple overlapping emitters in a single frame with a wide Gaussian envelope, resulting in an unusually high sigma value. By using the novel SiMoNa interface these defects could be efficiently detected.

Data filtering. After the identification of possible faulty localizations as described in the aforementioned scenario, a table of all localization parameters (Figure 2.9, top right) allows setting upper and lower thresholds to filter localizations. In the given example, Figure 2.10D shows the two identified peaks. By applying filters, the final population (shown in yellow) that represents a single Gaussian-like shape can be recovered. Ultimately, a scatter plot of the sigma and the offset parameter (not shown) could reveal where the faulty localizations are located, for example in areas of high fluorophore density.

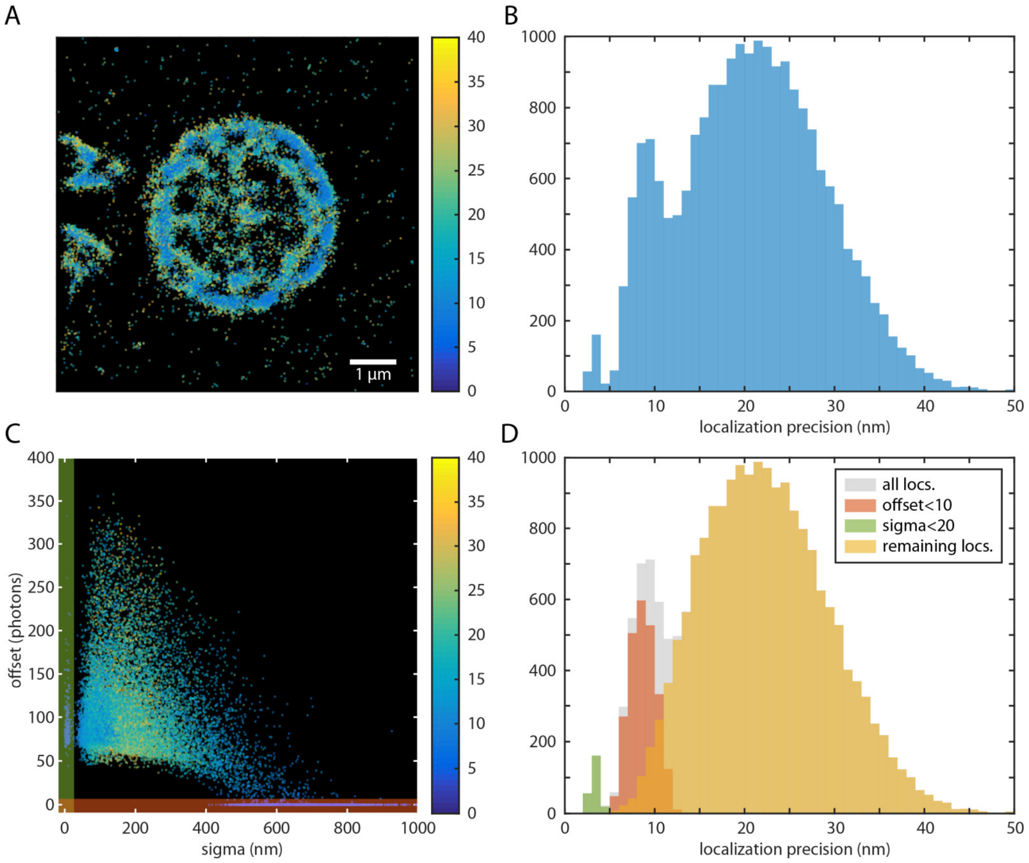


Figure 2.10 Spotting faulty localizations with SiMoNa A) Super-resolution scatter plot, colored by the localization precision shows inhomogeinities. B) The Histogram of the localization precision shows multiple populations. C) A three parameter plot between sigma and offset of the fitted Gaussian, colored by the localization precision reveals outliers in the dataset that can be filtered out by setting thresholds (indicated by the green and red shaded area). D) With the identified outliers (in green and red) removed, the localization precision now displays a single population (in yellow).

Data visualization. Additionally to the scatter plot introduced above, SiMoNa features multiple established super-resolution reconstruction algorithms suited for different tasks. Figure 2.11 will highlight the most important ones.

- Basic 2D histogram binning offers a representation with very short computation time (Figure 2.11A).
- The commonly used “Gaussian representation” represents each localization with a 2D Gaussian that features a width based on the localization precision. This super-resolution image representation can additionally be colored by the active parameter (Figure 2.11B and D).

- The density-based scatter plot can display the localizations color-coded according to a local density estimated by a running mean filter of a 2D histogram of the localizations (not shown).

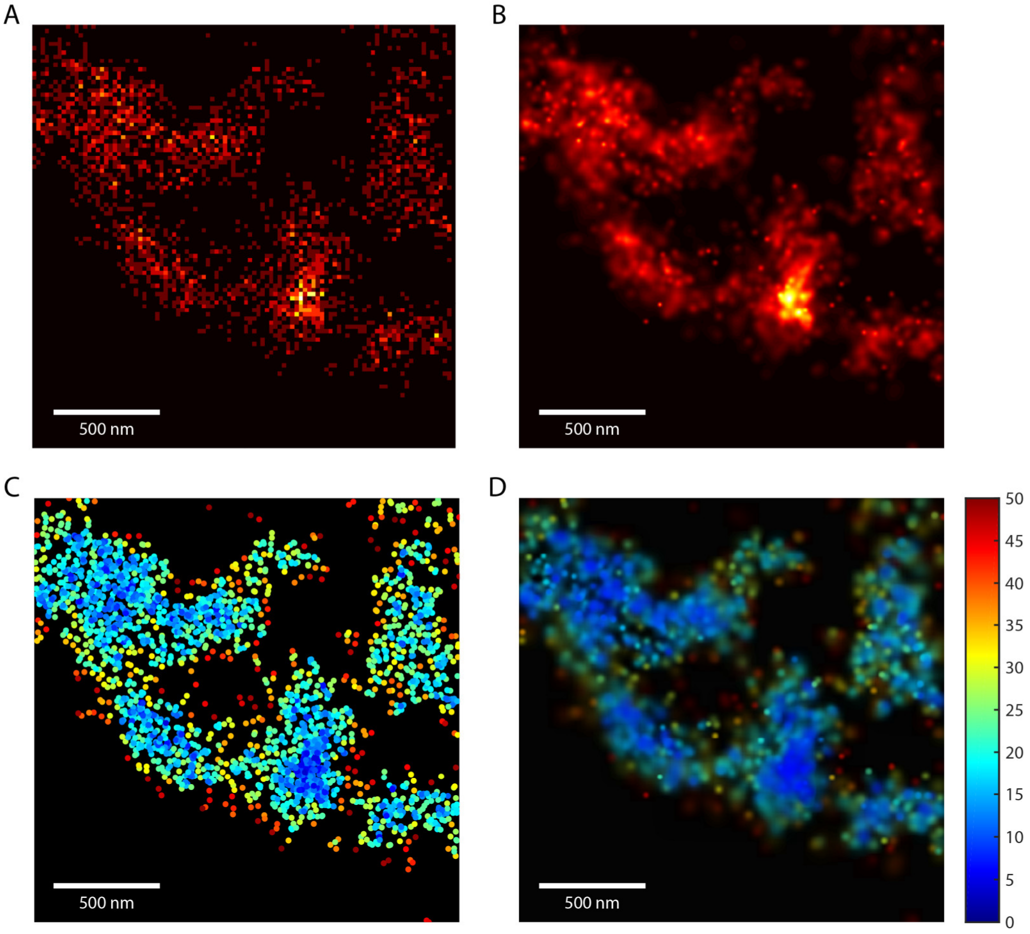


Figure 2.11 Comparison of different super-resolution reconstruction algorithms All plots show a zoom-in of Figure 2.10A. A) Histogram binning with a bin size of 20 nm. B) Gaussian representation of each localization with the localization precision as the width. C) Scatter plot colored by the nearest neighbor distance. D) Same as panel B), but colored by the nearest neighbor distance.

Post-processing. In addition to the data filtering and visualization methods, multiple algorithms concerning image resolution and clustering that have already been introduced in chapters 2.2 to 2.5 are implemented as well:

- Drift correction: For drift correction various software packages are available that are specifically designed to remove complex drift [43], [51], [70], [71]. Also ThunderSTORM has a powerful correlation-based drift correction built in. Therefore, only two features that complement these established techniques have been included in SiMoNa: A simple linear drift-correction allows the selection of a sub-region in the image that displays drift which is removed by a linear fit of the x and y position. Additionally, it is possible to import a drift file from an external source that describes the drift per frame which will be subtracted from the data.
- Fourier ring correlation for resolution estimates as in chapter 2.2.
- Nearest neighbor distance calculation for density analysis: The nearest neighbor distance can be used to efficiently remove sparse noise from super-resolution images.
- Calculation of Ripley's K, L and H function for identification of clustered datasets.
- Cluster identification and analysis toolbox: This includes finding clusters, separating them and analyzing localization patterns within single clusters. This workflow is used and explained in detail in chapter 3 where it is applied to DNA origami nanostructures to extract distances between individual fluorophore binding sites.

All these tools combined make SiMoNa a very valuable flexible toolbox that allows an initial and comprehensive exploration of SMLM datasets that focuses on quantitative analysis.

3 RESOLUTION OF THE SMLM SETUP TESTED WITH DNA ORIGAMI NANOSTRUCTURES

With the help of the data analysis framework established in chapter 2, the microscope setup resolution and SMLM acquisition routine was tested. An optimal "test sample" for this task would be a template of known geometry that can be modified with fluorophores of choice at distances between tens and hundreds of nanometer. It should be as rigid as possible and reproducibly producible. DNA origami has exactly these properties and is therefore an ideal length standard template to evaluate resolution in SMLM.

DNA origami has been initially described as the folding of a long, single strand DNA (scaffold strand) into pre-defined two and three-dimensional shapes using short single strand DNA (staple strands) [72]. These staple strands connect individual regions of the scaffold (by forming a double strand DNA with it) and give the structure its stability. DNA origami is synthesized autonomously by thermal annealing using a solution containing the scaffold and all the staple strands by cooling the mixture down from around 90 °C to ambient temperatures. Due to the availability of DNA synthesis technology, DNA origami can be designed in a huge variety of shapes ranging from simple helix bundles, planar sheets to complex 3d structures with functional elements.

In chapter 3.1 the specific DNA origami sample used for this work as well as the imaging and analysis routine will be introduced. Global resolution estimates as introduced in chapter 2.2 will be applied to the experimental data (chapter 3.2). Local resolution estimates tailored to the sample geometry of DNA origami are introduced in chapter 3.3 which enable accurate distance measurement between fluorophores anchored to DNA origami.

3.1 DNA origami as a length standard

The staple strands can be precisely modified with fluorophores to create inter-fluorophore distances of a few nanometers up to a micrometer. Thus, already established protocols to generate specific DNA origami shapes can be easily modified to create length standards for SMLM.

Sample design. As a starting point, a 2D triangular DNA origami structure [72], [73] has been chosen. Three staple strands at the bottom and two at the top are modified with biotin and the fluorophore Alexa647, respectively, (Figure 3.1A) to create fluorescence sources of exactly defined distances. The biotin allows immobilization of the sample on neutravidin coated glass slides. Three constructs were designed with inter-fluorophore distances of 96, 147, and 224 bp, resulting in 33, 50, and 76 nm distances between the pairs with the assumption of 0.34 nm/bp. In the following these constructs are just called "30, 50, or 76 nm origami", as shown in the scheme in Figure 3.1B. In order to verify the correct self-assembly, immobilized DNA origami was imaged using atomic force microscopy (AFM). The AFM images in Figure 3.1C&D show that at least 2/3 of the origami folded correctly into triangles. As the surface drying performed prior to the AFM measurements might additionally damage some origami triangles due to the resulting surface tension, it is safe to assume that the majority of triangles was intact during fluorescence imaging.

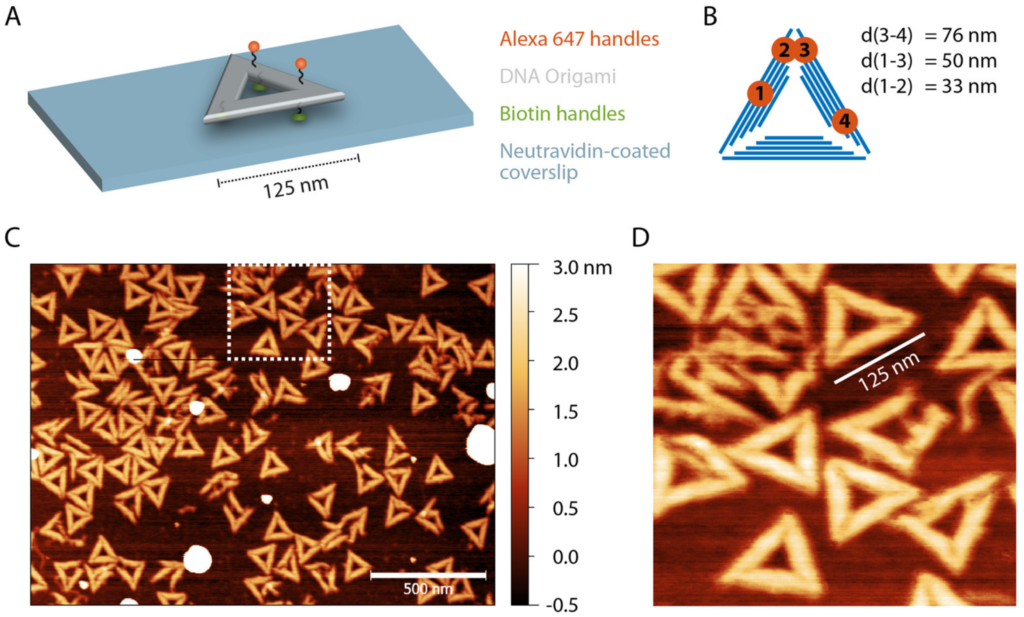


Figure 3.1 Sample design and AFM images of the DNA origami A) Schematic figure of a single DNA origami immobilized on a glass slide with two possible Alexa647 attachment sites displayed. B) The four available labeling position are displayed as well as the 3 selected pairs with their theoretical distances. C) AFM of the DNA origami on a dried mica surface shows the characteristic triangles. D) Higher resolution measurement of the area marked in C).

STORM imaging of immobilized DNA origami nanostructures. In order to induce the required blinking of the fluorophores, a buffered aqueous solution supplied with cysteamine hydrochloride was applied (see methods chapter 8.5 for details). Using TIRF microscopy, the Alexa647 fluorophores, bound to the immobilized origami nanostructures, have been imaged using a high excitation power (647 nm, 100 mW) and variable activation power (405 nm) to adjust the localization count. The resulting SMLM movie was analyzed by ThunderSTORM which gives direct access to the molecule list with several localization parameters – most importantly the x-y-position and the frame. The final super-resolution image shows a homogeneous distribution of DNA origami nanostructures over the whole field of view (Figure 3.2C) indicating that no aggregation of the sample took place. Pairs of fluorescent dots indicate a correct self-assembly of the nanostructure (highlighted as blue rectangles in Figure 3.2C).

Drift correction. As can be seen in Figure 3.2A, a single DNA origami gives rise to a smeared out distribution both in the widefield image and the super-resolution image. This is due to an x-y-drift over the course of the experiment, caused by thermal fluctuations. In this particular case, a drift of 443 nm over the whole experiment of 5000 frames (at 10 frames per second) has been observed, resulting in

53 nm drift per minute. In order to correct for that drift, a global cross-correlation-based drift correction as introduced in chapter 2.1 was used. The resulting cross-correlation calculated by ThunderSTORM was very stable (i.e. the cross-correlation peaks could be fitted with high precision) and the detected drift appeared almost linear. An exemplary curve of a DNA origami measurement is shown in Figure 2.1 (chapter 2.1). The corrected localizations result in a circular intensity distribution in the widefield image and reveal two clusters in the super-resolution image (Figure 3.2B) that can be further processed.

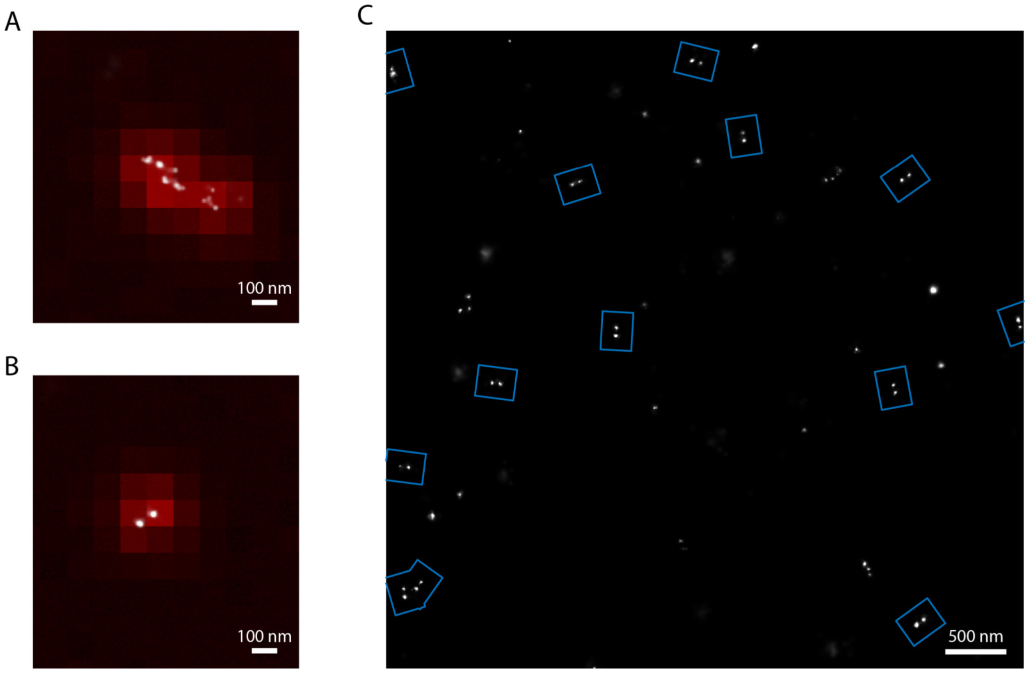


Figure 3.2 STORM images of a 76 nm origami A) Superimposed widefield maximum intensity projection (red) and super-resolution image (white) before drift correction. B) The same area as in A) after drift correction. C) Larger field of view of a drift-corrected super-resolution image. Double clusters that indicate intact origamis are highlighted in blue.

3.2 Global resolution estimates

As an initial step to test the resolution of the SMLM setup the complete dataset (i.e. all recorded localizations) was used as the data source and the localization parameter distributions were closely inspected. Here, the parameters intensity I, sigma σ and the localization precision Δx are of special interest. Figure 3.3 displays histograms of these localization parameters for the 76 nm origami sample. The 50 nm and 33 nm origami samples resulted in very similar distributions. The following checklist assures a high-resolution dataset:

- Sufficient localizations per frame are needed to form a dense super-resolution image. With a median of 37 localizations per frame (Figure 3.3A), and 10 frames per second, a dense image is formed within two minutes.
- Each localization has to have a high photon count to allow precise spatial determination (Figure 3.3B), which will result in a high localization precision.
- The width represented by the standard deviation σ of the fitted Gaussian profile for each localization has to be within a reasonable window. For this experiment a median σ of 124 nm (Figure 3.3C) was measured. The theoretical window can be calculated with the Rayleigh-criterion for the full-width half-maximum (FWHM): $FWHM = 2 \cdot \sqrt{2 \cdot \ln 2} \cdot \sigma = 0.61 \cdot \lambda / NA$. In this case, λ reflects the emission window of the microscope of 663 to 738 nm while the numerical aperture NA equals to 1.49 for the 100x oil objective used in this experiment. With these parameters, a theoretical window of 115 to 128 nm can be calculated, which matches the experimental median of 124 nm. By using the emission spectrum of Alexa647 as a weighting factor for the emission window, a theoretical mean emission wavelength of 687 nm is calculated that corresponds to a σ of 119 nm.

This checklist is on the one hand helpful to identify, quantify and compare the resolution and quality of the images and on the other hand also allows to filter out faulty localizations by visualizing the distributions of each parameter. These should consist of one major population – which is the case in the shown experiment.

Finally, the high photon count directly results in a high localization precision Δx – as calculated by the Thompson Webb formula introduced in chapter 2.2 – that can be used as a first resolution estimate [53]. For the DNA origami a mean localization precision $\langle \Delta x \rangle$ of 13.4 nm (Figure 3.3D) could be achieved which corresponds well with other STORM studies [33].

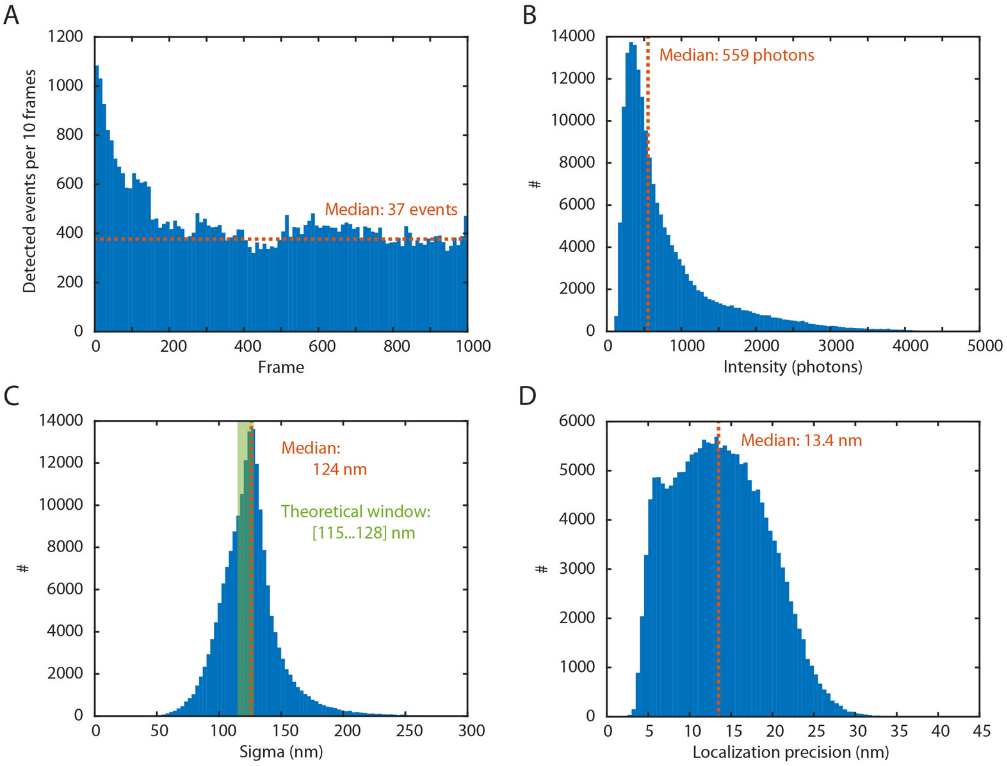


Figure 3.3 Localization parameter for the 76 nm origami. A) The histogram for detected localizations per frame shows fast bleaching within the first 200 frames that stabilizes at around 40 events per frame. B) The photon intensity per localization shows a clear peak with a median of 559 photons. C) The width (sigma σ) of the fitted Gaussian distributions display a peak at 124 nm that falls well into the theoretical window indicated in green. D) The localization precision shows a single population with a median of 13.4 nm.

In order to estimate whether the Nyquist theorem (see chapter 2.2) is satisfied, the density of the localizations was evaluated. A nearest neighbor distribution can be calculated by assigning to each localization the distance to its nearest neighboring localization. For a densely labeled sample, the average nearest neighbor distance has to be below the average localization precision $\langle \Delta x \rangle$. The 76 nm origami yields a median nearest neighbor distance of 2.6 nm (Figure 3.4A), which is clearly below the localization precision of 13.4 nm. This already indicates a dense labeling. Furthermore, Fourier ring correlation can build up on this to estimate a resolution threshold as described in chapter 2.2. The resulting resolution curve with an estimate of 25.3 nm is shown in Figure 3.4B. As this values lies below the smallest distance that has to be measured (i.e. the 33 nm origami sample), it is assured that a gap between two clusters spaced above 25.3 nm is indeed a gap and not an artifact due to sparse labeling or insufficient localization precision.

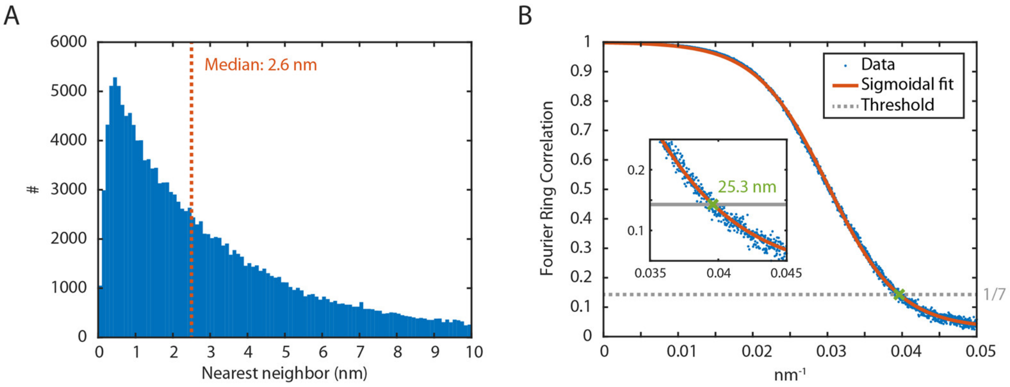


Figure 3.4 Nearest neighbor and Fourier ring correlation. A) The nearest neighbor distribution shows a very dense labeling with a median distance of 2.6 nm. B) The Fourier ring correlation gives an estimated resolution of 25.3 nm which is displayed as the intersection of the correlation function with the resolution criteria of 1/7.

As a final check, Ripley's H-function is calculated, which evaluates local point densities for radii r compared to a random distribution (as derived in chapter 2.4). Therefore, a value above zero for a radius r corresponds to clustering at those distances. For the DNA origami sample, the $H(r)$ curve is above zero for all distances, reaching its plateau at around 100 nm (Figure 3.5), which corresponds to the average cluster dimension of the 76 nm origami and therefore supports the assumption of a properly imaged SMLM dataset.

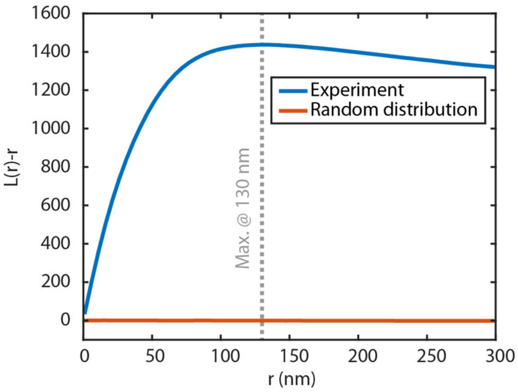


Figure 3.5 Clustering determination via Ripley's H function The random distribution (red) is expected to stay at zero for all explored distances r. The experiment (blue) shows significant clustering, reaching its plateau at around 100 nm.

3.3 Local resolution estimates

Compared to the aforementioned global resolution estimates that are performed on an entire SMLM movie consisting of hundreds of imaged DNA origamis, it is also possible to use local resolution estimates applied directly to individual DNA origamis to extract quantitative inter-fluorophore distances.

Line scanning. A straight forward approach to measure distances locally in a region of interest is the so-called "line scan" along an axis covering the structure to be measured. By comparing the diffraction-limited image with the reconstructed super-resolution image, it is obvious that SMLM allows to distinguish two fluorophore clusters for the DNA origami (Figure 3.2B & Figure 3.6A). By scanning across these two clusters, a peak-to-peak distance can be extracted by fitting the intensity distribution with a double Gaussian peak. An exemplary double peak for the 76 nm DNA origami is shown in Figure 3.6B, yielding a distance of 63.3 nm. As the choice of the line position is arbitrary, this method may introduce a big error as can be seen from the difference of more than 10 nm between experiment and theory. Additionally, this method is not based on raw localizations but a reconstructed super-resolution image. Therefore, the exact process of super-resolution image reconstruction will influence the final result. An unbiased and reproducible cluster analysis method will be introduced now.

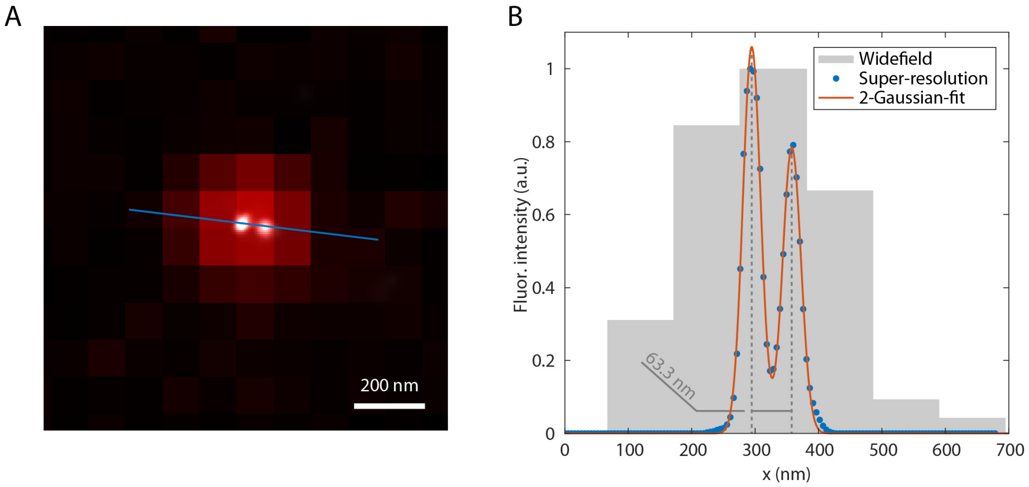


Figure 3.6 Comparison of widefield and super-resolution image for a 76 nm origami A) Overlaid widefield maximum intensity projection (red) and super-resolution image (white) with the line scan depicted in blue. B) Fluorescence intensity along the line scan. Pixel intensities from the widefield image in gray, super-resolution intensity in blue and the double Gaussian fit in red which yields a peak distance of 63.3 nm.

Unbiased cluster selection and analysis. While it is possible to select clusters visually by hand as shown for line scanning, an unbiased automated method is highly desired and helps achieving a large statistic. Although cluster algorithms like DBSCAN could be used, a custom routine optimized for the (generally known) DNA origami geometry will perform superior. Optimally, each selected DNA origami would consist of a cluster pair with a sufficient localization count where both fluorophores are emitting free of distortion, drift, noise or other imaging artifacts. Therefore, a selection routine has been established that uses raw localizations (after drift correction) to find clusters in the DNA origami dataset that can be further analyzed individually to extract the desired inter cluster distances:

1. To remove noise represented by randomly distributed localizations, a nearest neighbor filter is applied, excluding localizations with a nearest neighbor above 100 nm.
2. Single DNA origami clusters are identified by binning the localizations with a pixel size of 50 nm and applying an intensity threshold of 10 events per pixel. The resulting binary image is dilated with a square structuring element of the width 3 to include bordering regions around the clusters.
3. Clusters with less than 50 localizations are filtered out as an accurate distance measurement would not be possible.
4. Clusters whose diameter is too small (corresponding to a single very confined emitter) or too big (which are mostly multiple origamis close to each other) are filtered out by applying a two component k-means clustering and removing clusters with centroid distances below 20 or above 120 nm.
5. To further remove remaining clusters consisting of only a single peak (i.e. with only a single fluorophore), an ellipticity filter is applied. The ellipticity can be calculated by fitting an elliptical single component Gaussian to the clusters. By filtering out clusters with an ellipticity lower than 1.5, only elliptical distributions consisting of two peaks are selected.

The remaining clusters can then be analyzed automatically to extract the fluorophore distances.

Selection of cluster distance methods. In order to unravel the presence of two underlying sub-clusters spaced apart at a certain distance, the following methods are used and compared in this work: k-means clustering (KMC), expectation maximization algorithm for a Gaussian mixture distribution (GMD) and the novel PairRice fit (as introduced in chapters 2.4 and 2.6). All three methods were applied to all three origami samples to compare their ability to reproduce the correct distances. Figure 3.7A shows the raw localizations for a single cluster as well as the two species found by KMC and the probability density function for GMD. In Figure 3.7B, the pairwise distance distribution displays two populations that can be fitted with the PairRice fit, yielding a cluster distance of 79.2 nm.

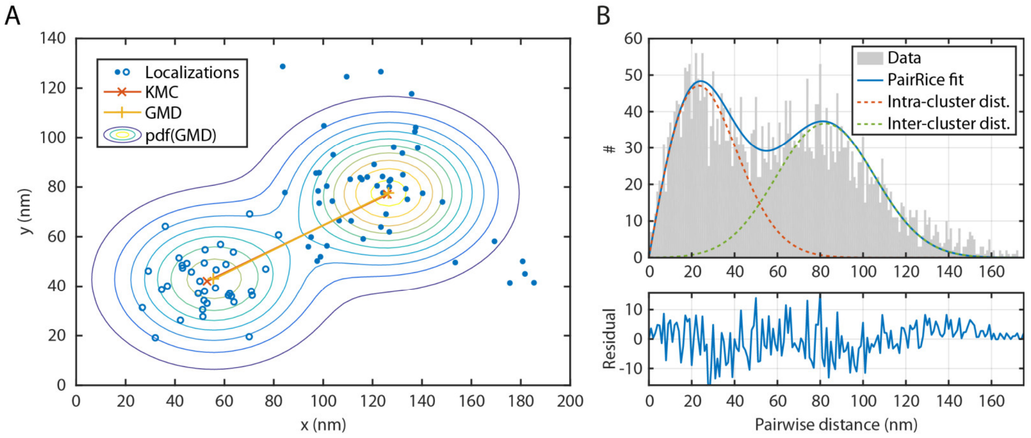


Figure 3.7 Length measurements in a 76 nm origami A) Raw localizations of a single DNA origami. Light and dark blue circles represent the two species detected by KMC. The contour shows the probability density function of the GMD algorithm from low (blue) to high (yellow). The resulting distances between the cluster centers are depicted by a red and orange line. B) The pairwise distance distribution of the localizations shown in A). The PairRice fit results in a cluster distance of 79.2 nm. The two underlying components (representing intra- and inter-cluster distances) are shown in red and green, respectively. The residual plot below supports the quality of the fit by displaying a homogeneous distribution over the whole distance.

Distance measurements on DNA origami. The resulting statistics for the whole dataset of all three origami samples are shown as box plots in Figure 3.8A with the standard deviations shown in panel B. For the 76 nm sample, all three fitting methods perform equally well. As 76 nm is the largest distance and therefore the easiest to fit, this is not surprising. The median distance is always around 68 nm with a standard deviation of 9 nm and therefore underestimating the expected distance of 76 nm, which could be caused by a slight bend of the DNA origami that reduces the distance between the fluorophores.

For the 50 nm sample, the KMC and GMD methods showed outliers that increase their standard deviation beyond 10 nm as compared to PairRice which yields a standard deviation of 8 nm. With around 47 nm the results fot the median is again similar in all methods and is very close to the expected distance. An increased inner quartile range indicates a slightly broader distance distribution. Since the fluorophores are anchored on different arms of the DNA origami triangle for this sample, it is reasonable to assume an increased flexibility between them compared to the other samples, where both fluorophores reside on the same arm of the triangle.

The 33 nm origami shows the biggest number of outliers for KMC and GMD while PairRice performs stably with a standard deviation at 5 nm. Nevertheless, the average determined distance is 35 nm for all methods, which agrees well within error with the 33 nm design.

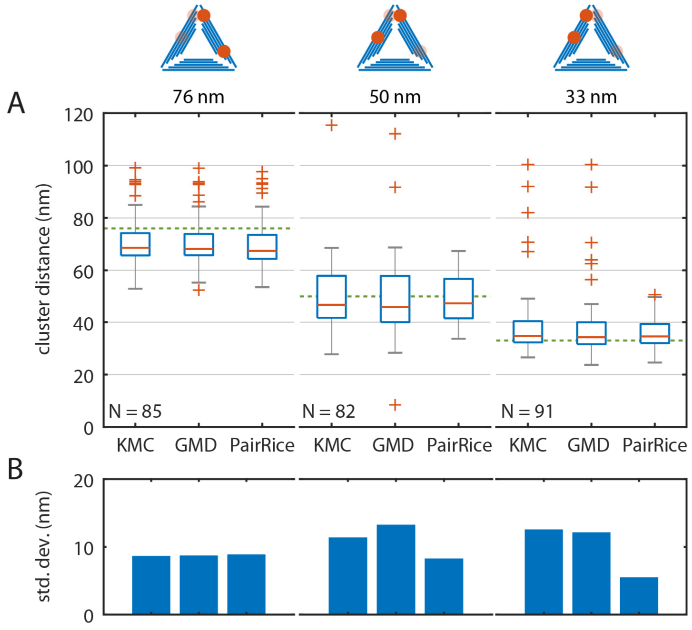


Figure 3.8 Evaluation of different fitting methods for different origami samples Overview of the three tested samples and the resulting distance distribution of all selected clusters per sample and per fitting method. A) The boxplots display the median (red line), the first and third quartiles (blue box), the 1.5 interquartile range (gray whisker), and the outliers (red crosses). The theoretical distance is plotted in green. The number of measured origami N is noted below. B) Standard deviations of above distributions.

Most outliers of both KMC and GMD appear to be at larger distances and stem from the fact that the algorithms identified few strayed localizations as a single small cluster and all remaining localizations as the second cluster. Therefore the distance between those clusters appears to be bigger. Since both methods find their optimal solution iteratively, they can get stuck in a local minimum and are dependent on the starting conditions (in this case – the location, width and distance of the two clusters) which are just “reasonable guesses”. In order to improve the quality of the distance measurements, it is possible to a) run the algorithm multiple times using different starting conditions to increase the chance of finding the global minima instead of a local minimum b) check the width and number of localizations per cluster to spot small or faulty clusters or c) add a weighting factor for outliers or points at the border of a cluster to reduce the risk of identifying these strayed localizations as a single cluster. The newly developed PairRice algorithm performed superior without major modifications and has thus been used in the following. Therefore an extensive optimization of the other two algorithms has not been undertaken.

Distance verification using simulated data. In order to verify and explore the different fitting methods in a controlled sample environment, they have been applied to simulated data of clusters spaced apart at different distances (Figure 3.9). Two Gaussian distributions of localizations spaced apart at 30, 40, 50 and 60 nm with a width (standard deviation) of 15 nm have been assumed. A total of 100 cluster pairs with 25 localizations in each cluster have been simulated, which falls in the range of the experimental data. Naturally, all methods perform better for bigger cluster distances as the cluster overlap is less. The KMC method constantly overestimates the simulated distance as it splits the cluster into two parts. Since this is only accurate if the two clusters are not overlapping, the calculated distance will always be larger than the simulation.

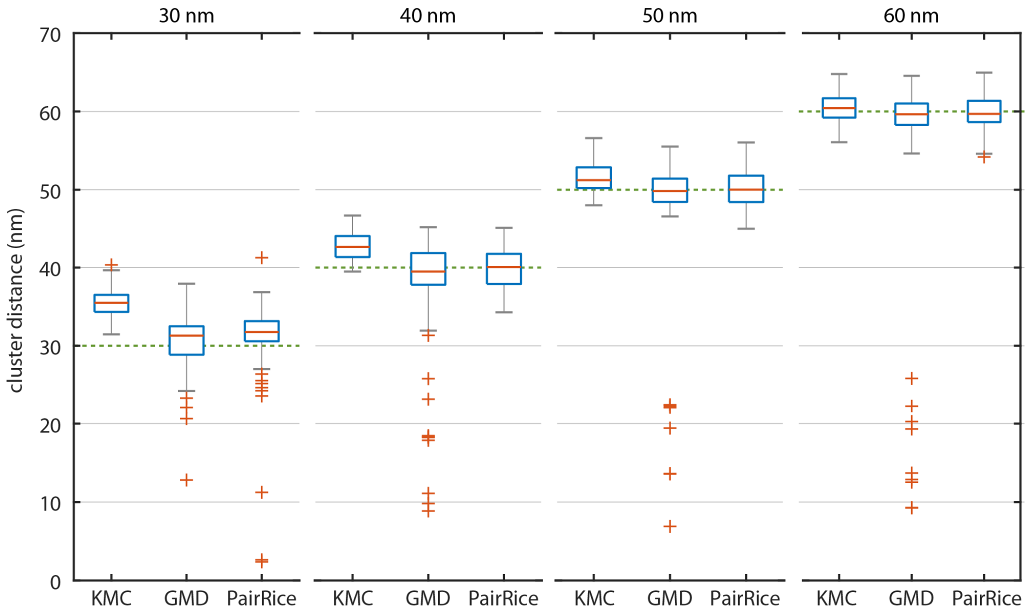


Figure 3.9 Evaluation of different fitting methods for simulated data Overview of multiple cluster distances and the resulting distance distribution per sample and per fitting method. Each simulation consists of 100 cluster pairs with 25 events in each cluster with a width of 15 nm. The boxplots display the median (red line), the first and third quartiles (blue box), the 1.5 interquartile range (gray whisker), and the outliers (red crosses). The simulated distance is plotted in green.

The GMD method on the other hand is not affected by this problem and therefore the distance evaluations are much more accurate on average. As in the case for DNA origami, outliers are the biggest problem for this method. Since stray localizations or noise are not included in the simulation, the outliers do not tend towards larger distances as seen for the DNA origami but to smaller distances. Again, the PairRice algorithm performed stable and superior, with a mean and median closer to the correct value than KMC while showing only few outliers for the smallest fluorophore distance of 30 nm.

3.4 Conclusion

It could be shown that DNA origami is perfectly suited to test the resolution of a SMLM setup by measuring the distance between two fluorophores. Triangular DNA origami nanostructures could be assembled reproducibly as AFM measurements confirmed.

SMLM could be performed with an average localization precision of 14.3 nm. An FRC resolution estimate confirmed the possibility to resolve spatial features down to 25.3 nm.

After establishing a cluster selection routine, only clusters containing correctly folded and labeled DNA origami structures with inter-fluorophore distances of 33, 50, and 76 nm were analyzed in an automated and unbiased way. This resulted in more than 80 analyzed structures per experiment giving adequate statistics.

A comparison between the novel PairRice algorithm with the two established clustering methods KMC and GMD showed that all methods are suited for the task of extracting fluorophore distances down to 33 nm, while the PairRice algorithm performed superior with regards to accuracy and precision. Furthermore, bending and flexibility of the DNA origami template could be observed and quantified. Simulations confirmed that the PairRice fit which I developed is the tool of choice, not only for the presented DNA origami experiment, but for estimating distances between two Gaussian-distributed point-clouds in close proximity in general.

4 EVALUATION OF PHOTO-CONTROLLABLE FLUORESCENT PROTEINS FOR PALM IN DIATOMS

One main aim of this doctoral work was to explore the intricate patterns of biosilica-associated proteins in diatoms. With the established analysis workflow (chapter 2) and the evaluation of the microscope and data analysis performance (chapter 3) in hand, the stage was set. Imaging biosilica embedded proteins that are involved in biomineralization features a specific challenge as these proteins are mostly not accessible to antibodies as introduced in chapter 1.2. Hence, STORM imaging is not possible without first dissolving the biosilica and therewith disintegrating the live diatom cell. Therefore, PALM imaging which uses genetically encoded photo-controllable fluorescent proteins (PCFPs) fused to the protein of interest is the method of choice if the conservation of the biosilica framework is desired.

The first step in the development of a PALM imaging strategy for the protein of interest is to choose between a huge variety of photo-controllable fluorescent proteins (PCFPs), each with its own strengths and weaknesses with regard to certain applications [34], [74]–[76]. Therefore, the first part of this chapter is dedicated to the selection process of promising PCFP candidates under the specific conditions required by diatoms.

4.1 Selecting PCFPs to minimize interference with the diatom autofluorescence

Photo-controllable FPs can be classified into three groups: photo-activatable FPs that can undergo an activation from a non-fluorescent (dark) state to a fluorescent state, photo-convertible FPs that can be converted from one fluorescent state to another, and photo-switchable FPs that can not only be converted once but reversibly switched between a fluorescent and non-fluorescent state. All three groups are suitable for PALM imaging and thus for imaging in diatoms. An important inclusion criterion for a possible PCFP candidate is its monomericity as oligomerization of multimeric PCFPs will influence the target protein pattern and thus introduce artifacts. Therefore PCFPs like tdEos [77], a tandem dimer as well as Kaede [78], KFP1 [79] or IrisFP [80], which are tetramers, were excluded.

Table 3 shows an overview of currently available monomeric PCFPs. This list is by no means complete but includes the fluorophores most commonly used for PALM. The six selected PCFP candidates are highlighted in green and the cover excitation wavelengths from 488 nm to 561 nm (Figure 4.1A). This allows the use of multiple PCFPs in the same sample e.g. in co-localization studies. Moreover, PCFPs with different conversion mechanisms were included in order to explore how the embedding in the diatom biosilica affects the conversion ability.

From the group of photo-activatable FPs, PA-GFP [81] in the green range (488 nm excitation) and PATagRFP [82] and PAmCherry1 [83] in the orange range (561 nm excitation) were chosen. The photo-convertible probes Dendra2 [84] and mEOS3.2 [85] were used as they exhibit a high brightness and are among the most frequently used PCFPs. From the photo-switchable FPs, Dronpa [86] has been chosen since it promises the highest brightness of all FPs under consideration. Figure 4.1A shows the emission spectra of all six chosen PCFPs.

Two further interesting candidates were PS-CFP2 [87] and PSmOrange [88] since they are the only PCFPs that can emit in the deep blue (405 nm excitation) and deep red (647 nm excitation), respectively. However, preliminary experiments indicated that in case of PS-CFP2 the fluorescence was too dim (which might be caused by bleaching due to the UV excitation). Moreover, the fluorescence signal overlaps with the emission of chlorophyll c and fucoxanthin at around 450 nm (Figure 4.1B), which impairs *in vivo* screening. PSmOrange features a similar problem, with its fluorescence signal overlapping with chlorophyll a emission at 670 nm (Figure 4.1B).

Table 3 Overview of PCFPs The most common PCFPs used in SMLM and their key characteristics are listed. The six PCFPs used in this work are highlighted in green. The abbreviations denominate: Act. WL: activation wavelength in nm, Pre act. state: state before activation, Exc. max: excitation maximum in nm, Em. max: emission maximum in nm, QY: quantum yield, Ext. coeff: extinction coefficient (in Mol^{-1} cm^{-1}), Brightness: the product of the extinction coefficient and the quantum yield divided by 1000. Host organisms are the jellyfish *Aequorea Victoria* (JAV), the anemone *Discosoma stratia* (ADS), the octocoral *Dendronephthya* (OD), coral *Lobophyllia hemprichii* (CLH), octocoral *Clavularia* (OC), and the coral *Pectiniidae* (CP).

Fluorophore	**Act. WL (nm)**	**Pre act. state**	**Exc. max (nm)**	**Em. max. (nm)**	**QY**	**Ext. coeff. (Mol^{-1} cm^{-1})**	**Bright-ness**	**Host**
Photo-activatable								
PA-GFP [81]	405	Dark	504	517	0.79	17000	40	*JAV*
PATagRFP [82]	405	Dark	562	595	0.38	66000	75	*ADS*
PAmCherry1 [83]	405	Dark	564	595	0.46	18000	25	*ADS*
Photo-convertible								
Dendra2 [84]	405	488	553	573	0.55	35000	57	*OD*
mEos3.2 [85]	405	488	572	580	0.55	32200	53	*CLH*
mMaple3 [89]	405	488	566	583	0.56	30000	50	*OC*
PS-CFP2 [87]	405	405	490	511	0.23	47000	32	*JAV*
PSmOrange [88]	488	561	634	662	0.28	32700	27	*ADS*
Photo-switchable								
Dronpa [86]	405 & 488 (off)	ON	503	517	0.68	98000	248	*CP*
Padron [90]	488 & 405 (off)	OFF	503	522	0.64	43000	82	*CP*
Kohinoor [91]	488 & 405 (off)	ON	495	514	0.71	63000	133	*CP*
rsFastLime [90]	405 & 488 (off)	ON	496	518	0.77	39000	89	*CP*
rsCherry [92]	550 & 450 (off)	Dark (40%F)	572	610	0.02	80000	5	*ADS*
rsCherryRev [92]	450 & 550 (off)	Dark (10%F)	572	608	0.005	84000	1	*ADS*
rsEGFP [93]	405 & 488 (off)	ON	493	510	0.36	47000	50	*JAV*

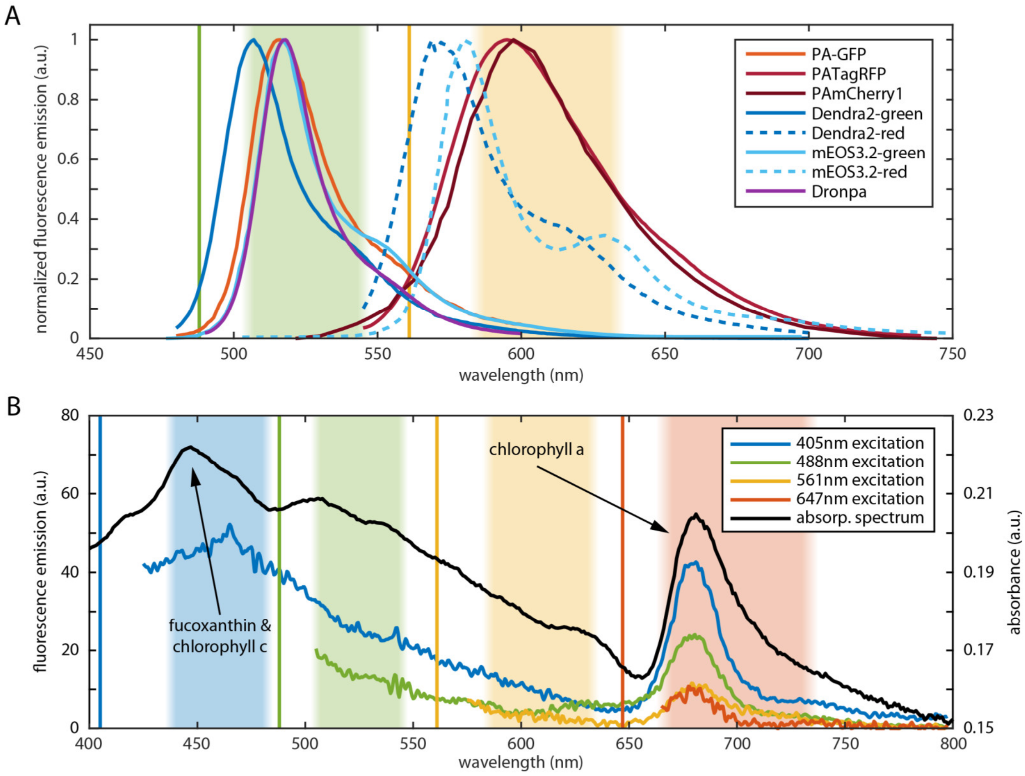


Figure 4.1 Emission spectra of selected PCFPs and live *T. pseudonana* Vertical lines represent the excitation laser lines. The shaded areas depict the respective detection windows. A) Normalized fluorescence emission of all PCFPs used throughout this work upon illumination with the appropriate excitation wavelength. B) Emission spectra for different excitation wavelengths as well as the absorption spectrum of live *T. pseudonana* cells (on a separate y-axis). Absorption for chlorophyll a peaks at around 670 nm, for chlorophyll c and fucoxanthin at around 450 nm [94].

4.2 Screening results for cytosolic and biosilica-embedded PCFPs

In a first set of experiments, the PCFPs were expressed in the cytosol of the diatom to test the photo-activation capabilities in a control situation without biosilica-embedment. For this purpose, the PCFP of choice was inserted into the expression vector pTpfcp for constitutive expression in the diatoms cytosol [5]. For details regarding the vectors, see methods chapter 8.3. The transformants were classified as successful when at least 25% of the screened clones showed fluorescence – if necessary after UV (405 nm) activation.

Figure 4.2 shows that all six PCFPs could be successfully activated or converted and showed the expected fluorescence when expressed in the cytosol. For PA-GFP and Dronpa conversion from the dark to the green state was detected. PATagRFP and

PAmCherry1 were activated from dark to orange fluorescence. Dendra2 and mEos3.2 were converted from green to orange fluorescence.

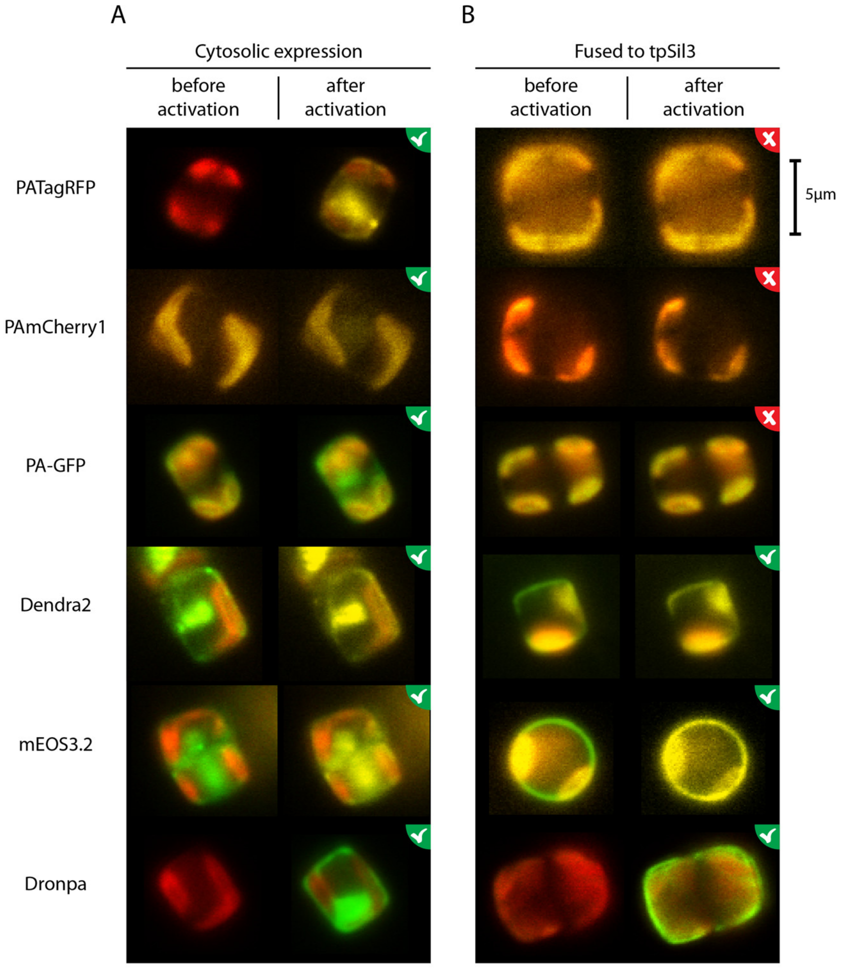


Figure 4.2 Screening of the fluorescent proteins and their activation capability
Epifluorescence images of *T. pseudonana* cells in side view. The activation was performed with the appropriate laser light as specified in Table 3. The colors represent the different excitation wavelengths 488 nm (green), 561 nm (yellow) and 647 nm (red). Chloroplasts are always visible in the red channel. A successful conversion is indicated by a green checkmark.

In a second set of experiments, the conversion of the PCFPs was tested when they were embedded in the diatom silica. For that purpose, fusion proteins of the PCFPs and Silaffin-3 (abbreviated with tpSil3 for the species *T. pseudonana*) were created. As tpSil3 is incorporated into all regions of biosilica when expressed under the control of the constitutive Tpfcp promoter [14], [20], [21] it is a suitable candidate to study the effects of biosilica-embedment on PCFPs. Plasmids encoding C-terminal fusion proteins of the six FPs and tpSil3 were constructed, transformed and expressed under the mentioned Tpfcp promoter (see methods chapter 8.3 for details).

In contrast to the cytosolic PCFP fusion proteins only three PCFPs remained photo-controllable when fused to tpSil3. Figure 4.2 shows that for all the photo-activatable FPs (PATagRFP, PAmCherry1 and PA-GFP), even strong UV illumination (1 second at 20 mW laser power) did not result in a fluorescence signal localized in the biosilica. By screening multiple cells per clone and at least 24 clones per transformation, it was ensured that the absence of fluorescence did not originate from a failed genetic transformation but indeed from a hindered photo-activation as a successful transformation would usually yield at least 10% of fluorescent clones.

The successful candidates, all the photo-convertible or photo-switchable FPs, could be converted from their green to orange state (Dendra2 & mEOS3.2) or from their dark to green state (Dronpa). The resulting fluorescence pattern resembles the cylindrical biosilica shape of the diatom very well. In order to study the conversion process in more detail, multiple cycles of UV illumination (1 second at 20 mW laser power) were alternated with epifluorescence imaging (Figure 4.3, top). Quantification of the fluorescence signal in a region containing the biosilica cell wall showed that the converted fluorescence saturates within a few cycles of UV illumination (Figure 4.3, bottom), indicating an unhindered conversion which is on a par with the cytosolic construct.

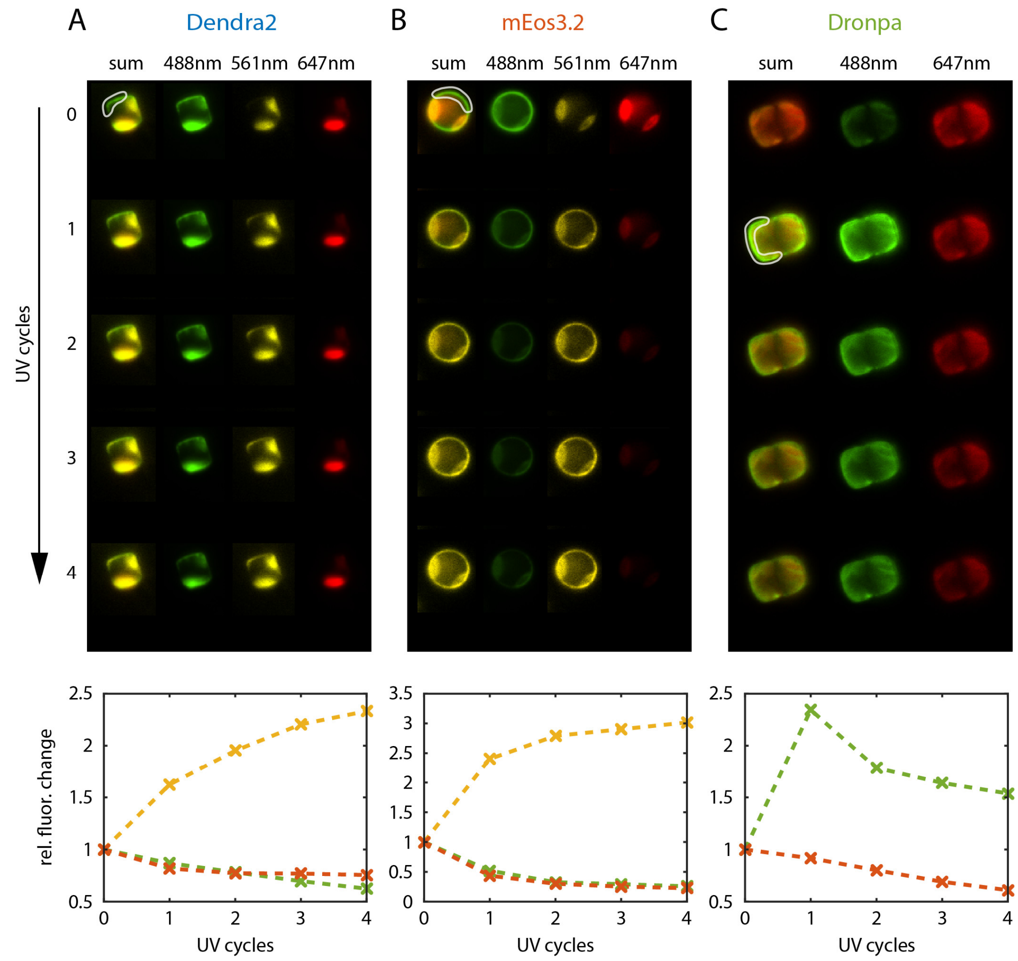
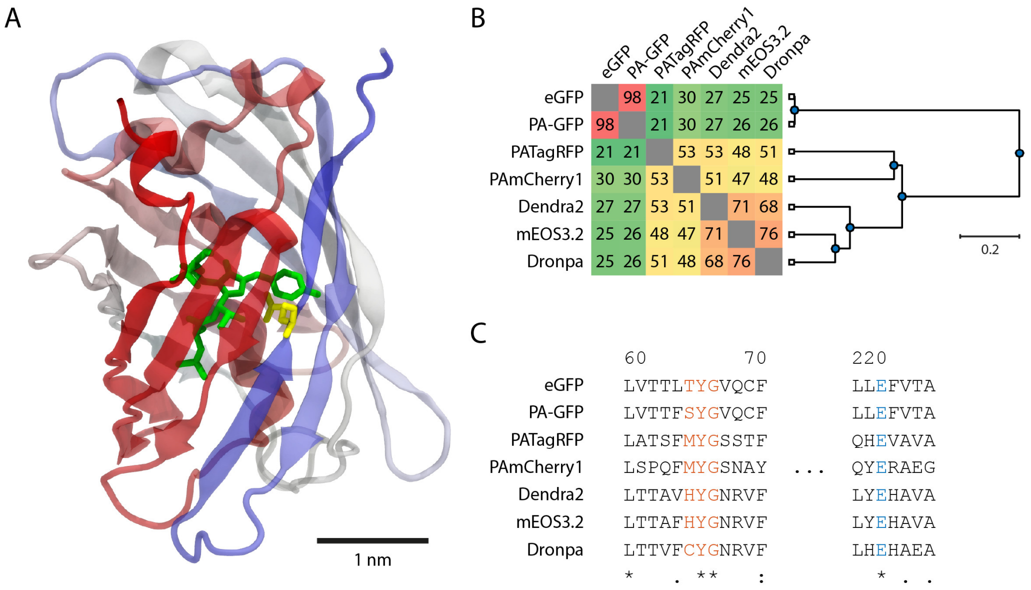


Figure 4.3 Photo-conversion statistics of active tpSil3 clones The three functional PCFPs Dendra2 (A), mEOS3.2 (B) and Dronpa (C) have been analyzed. The top panels show the individual imaging channels during conversion. Each UV cycle equals one second of 405 nm illumination at 20mW. In order to analyze the conversion quantitatively, a region containing the biosilica cell wall and overlapping as little as possible with chloroplasts was selected (marked with a white border). The intensity in this region was summed up. The relative intensity change with 488 nm, 561 nm and 647 nm excitation is plotted in the lower panels, showing the clear photo-conversion or -activation of the probes.

4.3 The underlying conversion mechanism

In order to understand the observed difference in photo-convertibility between the cytosolic and biosilica-embedded constructs, the protein structure as well as the molecular conversion mechanism has to be elucidated. All six tested PCFPs share the same general protein secondary and tertiary structure: a beta barrel structure consisting of eleven beta strands that shield the chromophore at residue 65-67 which is well-known from eGFP (Figure 4.4A). Additionally, the sequences share an average 46% sequence identity (ranging from 21 to 76%) as determined by multiple sequence alignment using Clustal Omega [95] and shown in the percent identity matrix in Figure 4.4B,. Common features like the *YG sequence for the chromophore and the glutamic acid at residue E222 that define the charged state of the chromophore and are necessary for its maturation [96], [97] are shared (Figure 4.4C). Therefore, the observed difference in photo-convertibility appeared surprising at first glance.



	eGFP	PA-GFP	PATagRFP	PAmCherry1	Dendra2	mEOS3.2	Dronpa
eGFP		98	21	30	27	25	25
PA-GFP	98		21	30	27	26	26
PATagRFP	21	21		53	53	48	51
PAmCherry1	30	30	53		51	47	48
Dendra2	27	27	53	51		71	68
mEOS3.2	25	26	48	47	71		76
Dronpa	25	26	51	48	68	76	

```
            60        70       220
eGFP        LVTTLTYGVQCF       LLEFVTA
PA-GFP      LVTTFSYGVQCF       LLEFVTA
PATagRFP    LATSFMYGSSTF       QHEVAVA
PAmCherry1  LSPQFMYGSNAY  ...  QYERAEG
Dendra2     LTTAVHYGNRVF       LYEHAVA
mEOS3.2     LTTAFHYGNRVF       LYEHAVA
Dronpa      LTTVFCYGNRVF       LHEHAEA
            *   . **   :         * . .
```

Figure 4.4 Sequence similarities for PCFPs A) Protein structure of eGFP (pdb:4EUL, [97]). Colored by residue from red to blue. The chromophore and Glu222 are highlighted in green and yellow, respectively. B) Percent identity matrix showing the sequence similarity color coded by value (from green to red). eGFP was added as a reference. To the right, a phylogenic tree created using UPGMA (Unweighted Pair Group Method Average) is shown. C) Sequence alignment of all PCFP sequences for the chromophore region (residue 65-67) and the region around Glu222, which is conserved. The last row shows the degree of conservancy as specified by the Clustal W file format.

In order to understand the differences in photo-convertibility of the biosilica embedded PCFPs, it is crucial to consider the underlying molecular conversion mechanism. In fact, the conversion mechanism of PA-GFP, PATagRFP and PAmCherry1 relies on a decarboxylation of Glu222 [98]–[100] (Figure 4.5A). This glutamic acid faces the outer surface of the β-barrel of the protein and is most likely partly exposed to the silica environment. Thus, it is conceivable that the surrounding silica environment impairs the decarboxylation. Interestingly, the three FPs that were converted successfully inside the biosilica do not require this decarboxylation. The conversion of Dendra2 and mEOS3.2 relies on a breakage of the polypeptide backbone near the chromophore [101] (Figure 4.5B) while Dronpa performs a cis-trans isomerization of the chromophore to switch between dark and fluorescent state [102] (Figure 4.5C). These three conversions occur in the center of the β-barrel, which is most likely protected from the surrounding biosilica.

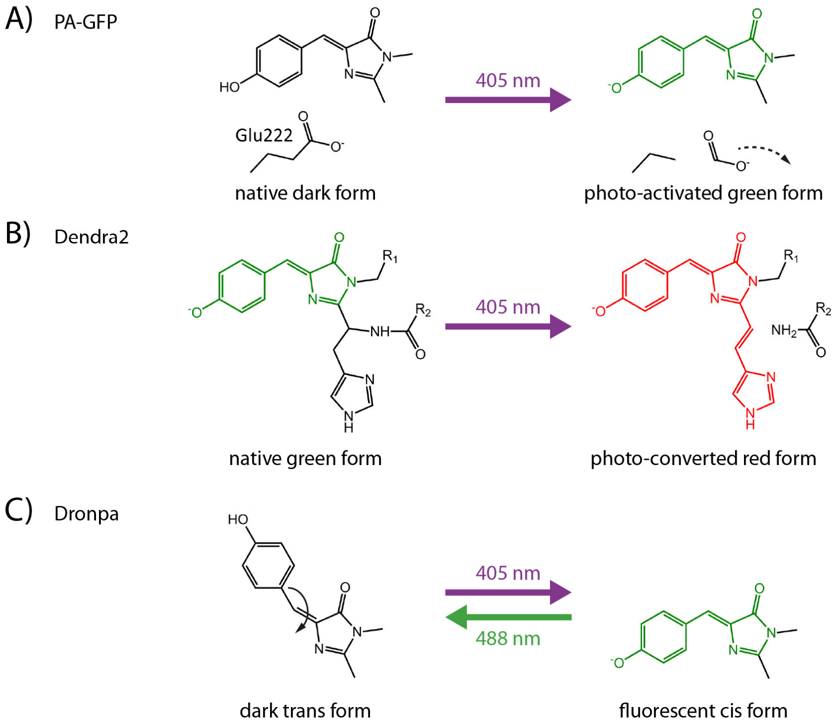


Figure 4.5 Types of chromophore conversion Only the matured and cyclized chromophore region is shown. A) Photo-activation of PA-GFP featuring deprotonation and decarboxylation. B) Photo-conversion of Dendra2 relies on deprotonation and polypeptide backbone cleavage. C) Photo-switching of Dronpa displays a cis-trans isomerization of the chromophore. The figures are adapted from [103], [104].

4.4 Conclusion

Six widely used monomeric PCFPs that are particularly suitable for the diatom organism due to minimal overlap with chloroplast autofluorescence have been screened for their suitability to be used for SMLM in diatoms. All PCFPs could be successfully introduced into the organism and were expressed and imaged in the cytosol.

In order to image proteins encased in the biosilica of diatoms, fusion proteins with the biosilica-embedded tpSil3 have been created. Surprisingly, only three of the six PCFPs (Dendra2, mEOS3.2 and Dronpa) were still convertible when fused to tpSil3. This phenomenon could be explained by the underlying conversion mechanism of these PCFPs: the conversion takes place in the interior of the FP near the chromophore and is thereby shielded from the biosilica environment.

With the discovery of multiple PCFPs that are convertible inside biosilica, the way is paved for PALM imaging of biosilica-embedded proteins like tpSil3, enabling biomineralization studies at high spatial resolution, as will be shown in chapter 5. Even multicolor PALM imaging is feasible by using Dronpa in conjunction with mEOS3.2 or Dendra2.

5 IMAGING THE SIL3 MESHWORK

In previous studies, tpSil3-GFP has already been localized using confocal fluorescence microscopy and showed incorporation into all regions of the biosilica [21]. However, a more detailed study exploring its distribution pattern that could contribute to the understanding of the role of tpSil3 in the biomineralization process of the diatom cell is still missing. Hence, one of the aims of this doctoral work was to study this protein pattern using SMLM. The increased resolution gained by SMLM will additionally allow the investigation of these patterns with unprecedented precision.

In the preceding chapter the successful identification of Dendra2, mEOS3.2 and Dronpa fusion constructs as suitable PALM probes for imaging biosilica embedded proteins and tpSil3 in particular was presented. This chapter presents the results of PALM imaging of these fusion constructs.

5.1 Analyzing protein layer thickness using tpSil3-Dendra2

Since a high signal to noise ratio is necessary to assure precise single fluorophore localization, all significant background fluorescence has to be eliminated. Like in all photosynthetic organisms chloroplasts are present in the diatom cell which have a wide autofluorescence spectrum (as shown in Figure 4.1B) that overlaps the fluorescence of the introduced FPs. Even the chosen FPs like Dendra2 that feature a minimal overlap with the chloroplast autofluorescence were not bright enough to yield a signal to noise ratio sufficient for single fluorophore localization in presence of the high background. Therefore, the chloroplasts needed to be removed by extraction of the cells with a detergent-based buffer retaining the pure biosilica skeleton and the embedded proteins. See chapter 8.3 for a detailed protocol. Since the fluorophores are shielded from the liquid buffer environment, they survived the relatively harsh extraction procedure [17].

For imaging of the tpSil3-Dendra2 and tpSil3-mEOS3.2 fusion proteins, a 488 nm excitation of 20 mW was sufficient to localize pre-converted FPs. For Dronpa, a continuous low (UV) conversion power of 8 mW was necessary to convert it from the dark ground state to the fluorescent state. After a relatively short recording time of approximately two minutes (1000 frames at 10 frames per second), the resulting movie was analyzed using ThunderSTORM and the localizations were visualized using the SiMoNa GUI. By imaging tpSil3-Dendra2 with the *T. pseudonana* diatom in valve view and the focal plane in the girdle band region, the thickness of the tpSil3 filament can be visualized very well, as illustrated by the scheme in Figure 5.1A and the reconstructed single-molecule localization image in Figure 5.2A. By performing multiple line scans perpendicular to the biosilica cylinder, an average full width at half maximum (FWHM) of 76.0 nm ± 4.6 nm (± S.D.) was determined for the protein layer. One of those line scans is depicted in Figure 5.2B.

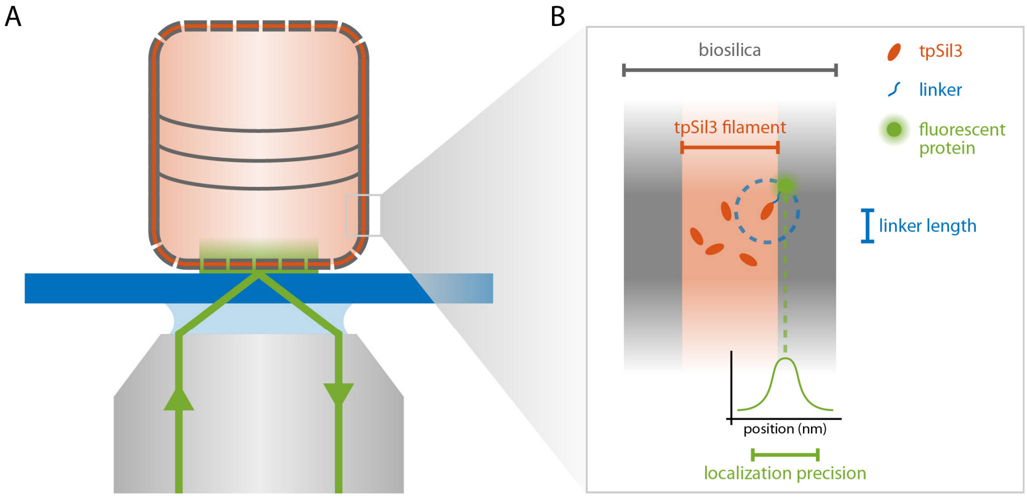


Figure 5.1 tpSil3 illumination scheme A) Illumination scheme for tpSil3 studies. The diatom is imaged in valve view with the focus being either near the surface for valve imaging or in the girdle band region for protein filament thickness measurements. Laser excitation was in total internal reflection mode. B) Zoom-in showing a scheme of the tpSil3 filament highlighting the different uncertainty contributions noted in the main text.

It has to be considered that the determined width of 76 nm in fact does not represent the actual filament thickness, but rather a convolution of this value with two additional components, consisting of linker length and localization precision (Figure 5.1B). Therefore, deconvolution helps unraveling the actual underlying filament thickness, which is assumed as a rectangular function of unknown width (Figure 5.2D). The following contributions have been considered:

- The localization precision per fluorophore is 28 nm. This corresponds to a FWHM of 66 nm assuming a Gaussian distribution (Figure 5.2E)
- The linker length (the distance between the chromophore of the PCFP and tpSil3) is around 3 nm. This corresponds to points on a sphere with a radius of 3 nm (Figure 5.2F)

Hereby, the 76 nm FWHM can be deconvoluted to an underlying thickness of 53 nm ± 3 nm, meaning that a 53 nm wide rectangular function adds up to the measured 76 nm wide Gaussia when convoluted with all the mentioned contributions. This value coincides with the values mentioned in scanning electron microscopy studies that reported an average thickness of the valve biosilica of 60 nm [105]. From these measurements, it can be concluded that tpSil3 does not form a singular flat protein layer which would result in a thickness of below 10 nm, but in fact spans almost the whole biosilica cell wall with its size of 60 nm as a denser meshwork. For comparison, the line scans of the epifluorescence image, which is limited in its resolution by the diffraction limit, yielded a FWHM of 510 nm ± 55 nm (Figure 5.2B), which only confirms that tpSil3 is part of the biosilica shell, but allows only limited conclusions on the actual filament thickness.

Additionally, a Fourier ring correlation was calculated to estimate the resolution [56] of the SMLM reconstruction. For the super-resolution image in Figure 5.2A, a resolution of 74.7 nm (Figure 5.2C) is achieved according to the Fourier ring correlation criteria, which corroborates the observed FWHM of 76 nm of tpSil3-Dendra2.

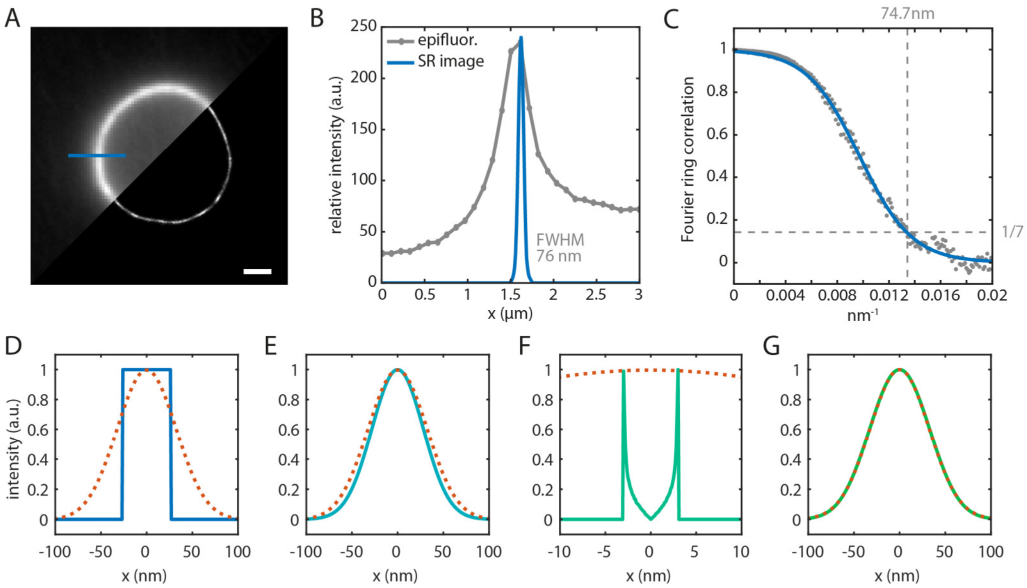


Figure 5.2 tpSil3-Dendra2 thickness measurements A) Comparison of an epifluorescence image and the reconstructed super-resolution image of tpSil3-Dendra2 with z-focus on the girdle band area of the diatom. The position of the line scan is highlighted. B) Line scan through the silica cell wall showing the fluorescence intensity profile for both imaging modalities. The FWHM of the PALM image is denoted. C) Fourier ring correlation for the super-resolution image shown in A) reveals an effective resolution estimate of 74.7 nm. D-G) Convolution process to estimate the underlying filament thickness based on the measured filament thickness of 76 nm (FWHM) in the super-resolution image of A)-B). The Gauss profile of the measured filament is displayed in red in all plots. D) Underlying filament with a thickness of 53 nm. E) The first convolution contribution: Fluorophore localization precision of 28 nm. F) The second convolution contribution: Linker length between protein and fluorophore of 3 nm. Assumed as points on a sphere of 3 nm radius. G) Final convolution of D) to F) resulting in the desired width of 76 nm.

5.2 Imaging the valve region using tpSil3

The intricate biosilica pattern of diatoms is most planar in the valve region due to the cylindrical diatom shape and can thus be best observed there. TIRF imaging was used to determine the locations of tpSil3-Dendra2, -mEOS3.2 and -Dronpa embedded in the biosilica. In epifluorescence images, only a continuous fluorescence distribution with inhomogeneous intensities over the valve region was observed (Figure 5.3A-C, left row). In the PALM images on the contrary clear fluorescent and non-fluorescent regions could be identified (Figure 5.3A-C, middle row).

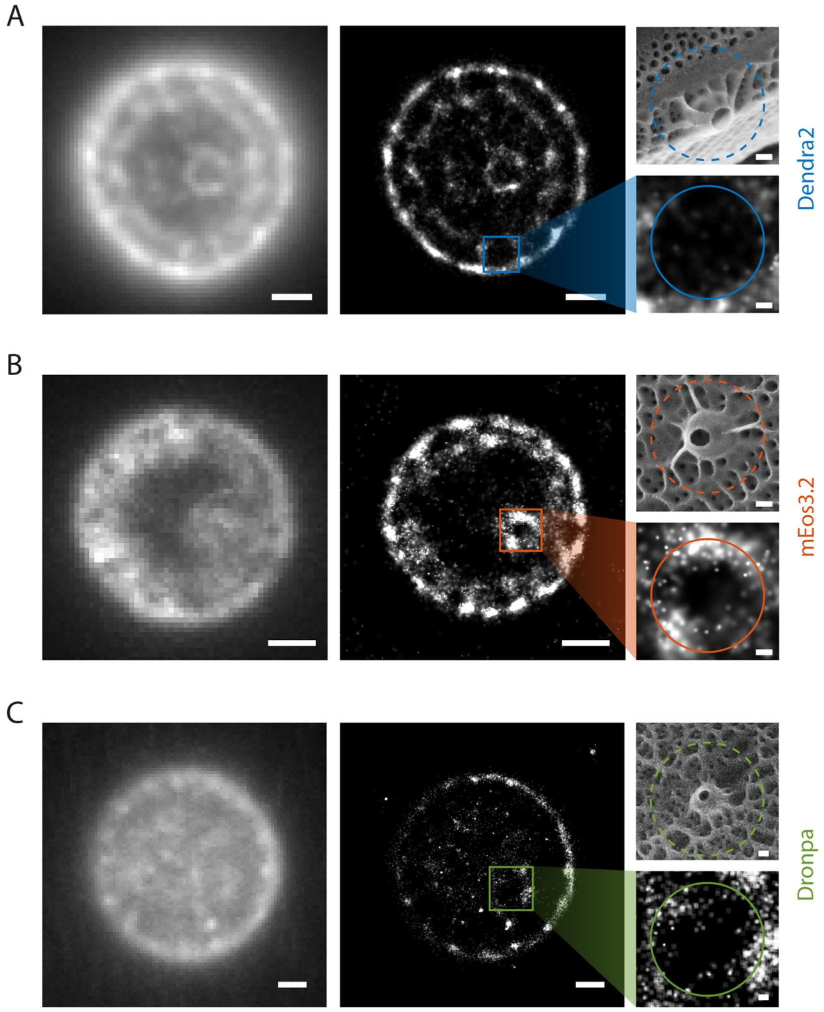


Figure 5.3 PALM of tpSil3 in the valve region Comparison of epifluorescence images and the reconstructed super-resolution image of A) Dendra2, B) mEOS3.2 and C) Dronpa fused to tpSil3 with z-focus on the valve region of the diatom. Enlarged details of the fultoportulae to the right. An SEM image in the same scaling that corresponds to a similar region is displayed for comparison. Circular areas of non-fluorescent gaps are highlighted for qualitative comparison. Scale bars are 1 µm, and 100 nm for the zoomed images.

Interestingly, the large circular non-fluorescent gaps correlate in number, size and position with the structural elements of the marginal as well as the central fultoportulae observed in wildtype *T. pseudonana* in SEM images (Figure 5.3A-C, SEM images). This suggests that the tpSil3 fusion constructs are not present within the external tubes of the fultoportulae which are approximately 100 nm in diameter, but rather correlate to the outer region of the fultoportula basal chamber which is a thicker and larger structure. This becomes apparent when looking at the inside of the valve biosilica via SEM (Figure 5.4). Furthermore, no distinct ridge-like patterns in the reconstructed super-resolution images could be observed, but a rather sparse localization on the central area of the valve, except for a high density around a central pore-like structure. The position and size of this pore-like structure correlates well with the central fultoportula observed in electron micrographs. Therefore, it is reasonable to assume that tpSil3 is a major component in the basal chamber of the fultoportulae and plays only a minor role in ridge formation. In order to further strengthen these assumptions, the localization parameters of the SMLM images will be analyzed in the following section.

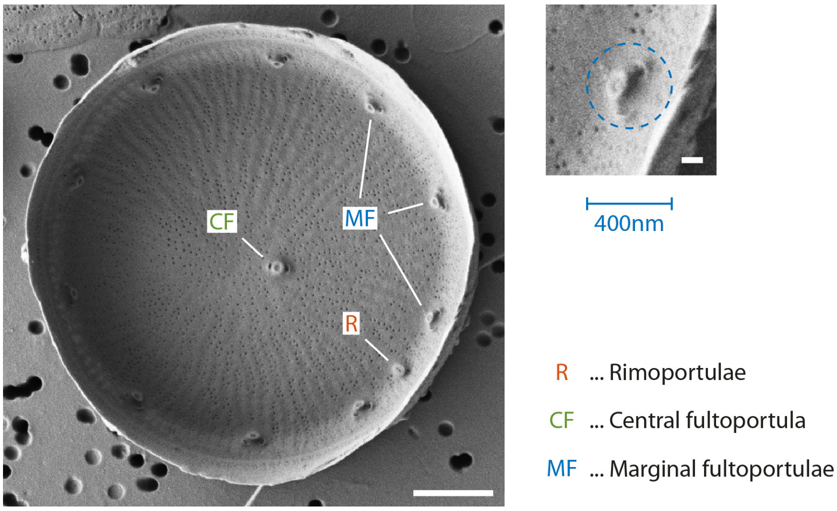


Figure 5.4 SEM image of the inside of isolated valve biosilica of *T. pseudonana*
The inside of the valve shows the thick basal chamber of the fultoportulae. A zoomed image shows a diameter estimate for the outer region of the fultoportulae basal chamber of around 400 nm. Scale bars are 1 µm and 100 nm for the zoomed image.

5.3 Resolution and localization parameters of tpSil3

In the valve regions, 17763 fluorophores (tpSil3-Dendra2), 14327 fluorophores (tpSil3-mEOS3.2) and 13716 fluorophores (tpSil3-Dronpa) were localized to form the corresponding super-resolution images shown in Figure 5.3. The average localization precision for Dendra2, mEOS3.2, and Dronpa is 28, 25, and 25 nm, respectively. The distribution in Figure 5.5A shows a typical single Gaussian shape. This is in good agreement with previously reported localization precisions for these fluorophores in different biological environments [31], [76]. Although there have been reports of light trapping [106] and lensing [107] effects of the nanostructured diatom frustules, a pronounced distortion of the far-field image leading to inferior fluorophore localization has not been observed, which indicates that silica-embedding does not affect imaging resolution with these fluorophores. The high amount of localized fluorophores for Dendra2 and mEOS3.2 (with a mean of 14 localizations per frame, see Figure 5.5B) and the localization precision support the hypothesis that the non-fluorescent areas are indeed gaps in the tpSil3 protein arrangement rather than artifacts generated by sparse image reconstruction. For tpSil3-Dronpa, the localization count per frame was considerably lower (5 versus 14). This leads to a longer acquisition time (2500 frames) until a similar super-resolution image is formed (Figure 5.3C). The lower localization count might originate from a lower expression of the PCFP fusion protein or from bleaching during acquisition due to the UV activation laser.

Further analysis of the nearest neighbor distribution clearly peaks under 18 nm for all SMLM images, thus indicating a dense protein pattern (Figure 5.5C) that satisfies the Nyquist Criterion [54] introduced in chapter 2.2.

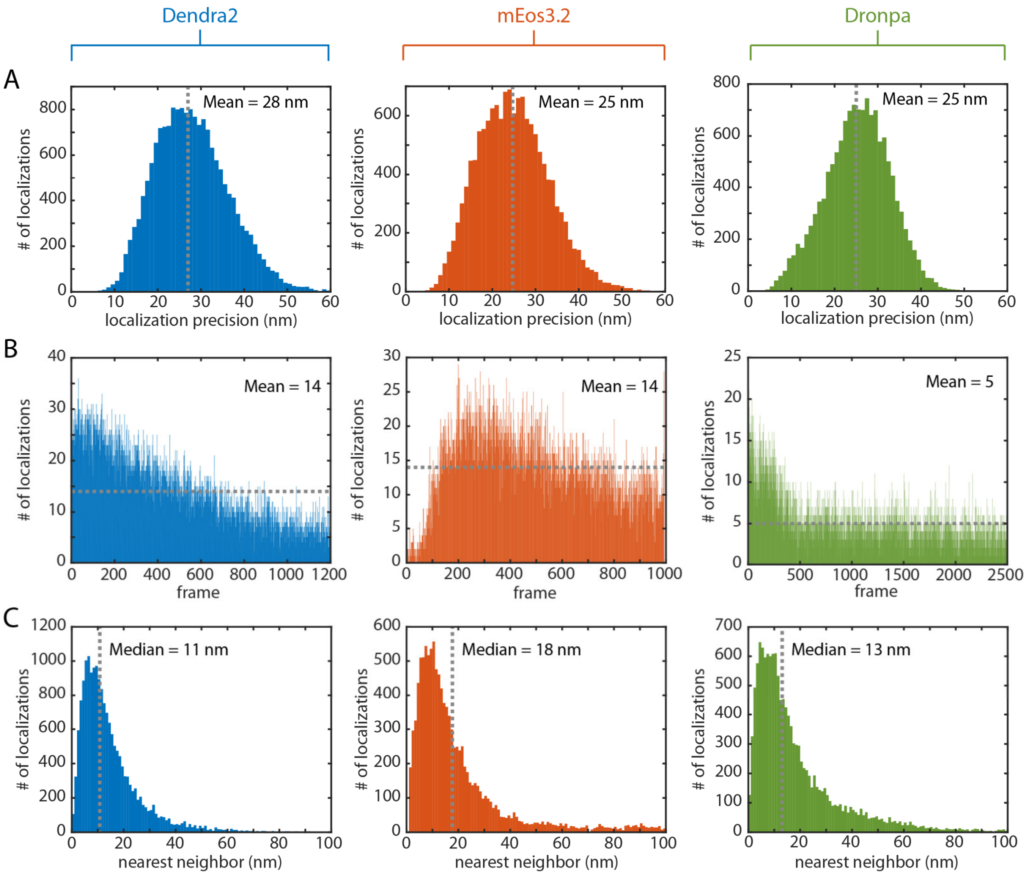


Figure 5.5 Localization parameter for tpSil3 fusion proteins Parameters of the reconstructed super-resolution image in Figure 5.3 A, B and C for Dendra2, mEOS3.2 and Dronpa, respectively. A) Localization precision of all localizations. B) Number of localized events per frame. C) Nearest neighbor distribution of all localized events.

5.4 Conclusion

By measuring the tpSil3 filament thickness using PALM imaging, it could be shown that SMLM allows insights into the biomineralization processes. It can complement and expand electron microscopy studies which only report on biosilica surface morphologies and are not able to visualize embedded proteins.

Furthermore, imaging in the valve region revealed correlations between the fluorescence patterns from super-resolution images and SEM studies imaging the biosilica surface. tpSil3 could be localized with high densities in the basal chambers of the fultoportulae, inferring an important role in formation of these biosilica structures.

Analysis of the SMLM localization parameters yielded localization precisions of 28 nm for silica-embedded tpSil3-Dendra2 and 25 nm for tpSil3-mEOS3.2 and tpSil3-Dronpa. This resolution is well beyond the diffraction limit and in line with other PALM studies [76]. Thus it has to be noted that the localization of biosilica embedded proteins has been studied with unprecedented resolution in this work. This ultimately will allow a deeper understanding of the molecular mechanisms that control protein-controlled silica biogenesis of diatoms.

With the availability of PALM imaging to study protein patterns in diatoms, it is possible to expand the established protocol to multi-color imaging which allows co-localization studies between different protein families. This will be shown for the relative position of different cingulin proteins in the following chapter 6.

6 DECIPHERING CINGULIN PATTERNS WITH CO-LOCALIZATION STUDIES

As already outlined in the introduction to this thesis, six girdle band-associated silaffin-like proteins have been discovered in 2011 by Scheffel *et al.* [17]. These proteins are a component of an insoluble silica-forming organic matrix: the microrings. This group of silaffin-like proteins has been termed cingulins. So far, conventional diffraction-limited fluorescence microscopy studies using cingulin-GFP fusion constructs identified the rough localization of these proteins in the girdle band area. However, the function of individual cingulins is still unknown. To investigate this was a main motivation for this chapter of the doctoral work.

It is conceivable that individual cingulins are not equally distributed within the microrings but perform a specific task at a specific location. In order to shed light on this question, co-localization studies at high spatial resolution are necessary. As a starting point, the relative position between two cingulins of different families, CinW2 and CinY2, will be examined.

6.1 A two-color cingulin construct for PALM-STORM

The most straightforward approach for a two-color super-resolution construct is to co-transform two different PCFPs attached to two different cingulins for PALM imaging. Unfortunately, this strategy was not applicable here as all the tested PCFPs that were functional when embedded in biosilica emit in the green channel (488 nm excitation) before (mEOS3.2 & Dendra2) or after (Dronpa) photo-conversion. Therefore, the emission in this channel would overlap thus making it impossible to distinguish the two species. This lack of available PCFPs for this experiment required the combination of PALM with STORM by using an additional fluorescent probe that is excitable by 647 nm light. The fluorescent dye Alexa647 has been chosen as it performs superior in STORM imaging [38].

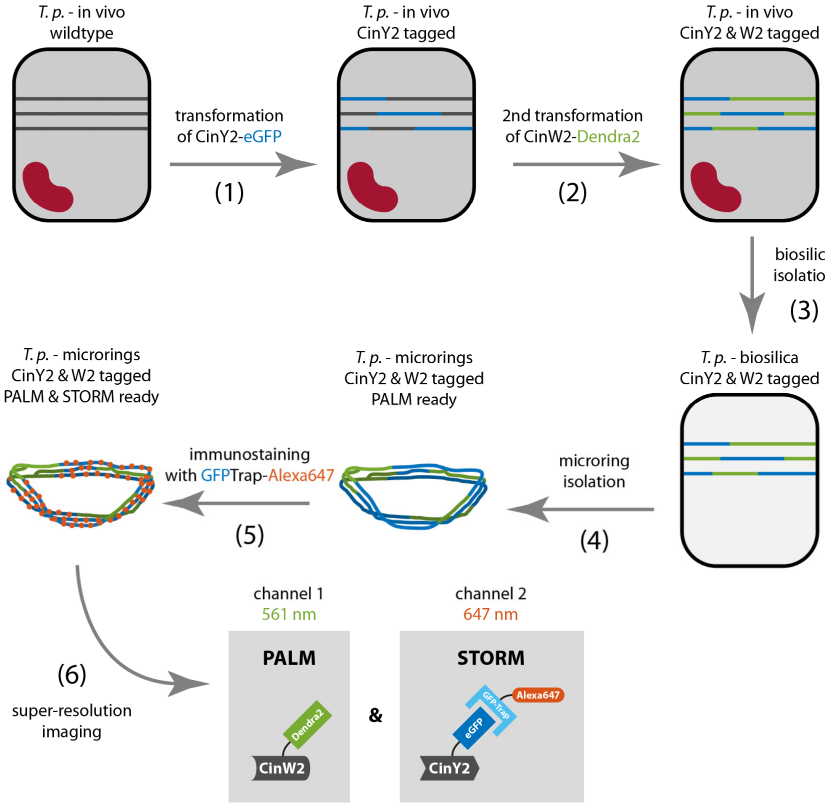


Figure 6.1 Transformation and imaging preparation workflow for the two-color cingulin construct The individual numbered steps are explained in detail in the text. The cell is represented in gray. A live cell is indicated by a red chloroplast. The two transformants, in this case CinY2-eGFP and CinW2-Dendra2, are shown in blue and green, respectively.

Figure 6.1 shows the workflow of the novel approach to create a diatom construct that can be used for a combined PALM and STORM measurement resulting in a two-color cingulin localization. The individual steps are described in detail in the following, while the numbers in brackets refer to the figure:

(1) As for the tpSil3 constructs, the first step was to create clones that feature a fusion construct between one of the 6 cingulins and the fluorescent protein of choice.

In order to assure natural expression, the native promoters for all cingulins were used. The first transformation involved introducing a CinY2 C-terminally tagged with eGFP (CinY2-eGFP) into the organism. The resulting clones can then be easily screened for fluorescence by excitation at488 nm.

(2) For the second transformation, CinW2-Dendra2 was chosen as the PALM probe, since Dendra2 performed superior when fused to tpSil3. It is important to introduce Dendra2 as the second step as its fluorescence signal may appear both in the 488 nm as well as in the 561 nm channel and might thus overlay the eGFP signal in the 488 nm channel. This is also the reason that a co-transformation of both CinY2-eGFP and CinW2-Dendra2 was not pursued as the the detection of CinY2-eEGFP in CinW2-Dendra2 positive clones would have been impaired.

(3) In order to enable the labeling of CinY2-eGFP with Alexa647 it is first necessary to isolate the biosilica by removing cytosolic components in a lysis step.

(4) To completely remove the shielding biosilica and make the cingulin accessible for the fluorescent STORM probe Alexa647, it was dissolved by an ammonium fluoride treatment. Since the cingulins are – in contrast to tpSil3 – part of an insoluble matrix, the remaining microrings containing the cingulins can be deposited on a surface. Poly-L-lysine coating of the glass slide supports this process.

(5) The following immunostaining was performed with a standard staining protocol using the so called GFPTrap® by ChromoTek, a single-domain antibody against GFP. These antibodies have been labeled with Alexa-647 via an NHS ester reaction. See the methods chapter 8.3 (“Biosilica isolation” and “Microring isolation and immunostaining”) for a detailed protocol of steps (3), (4), and (5).

(6) The final construct was imaged using variable illumination at 405 nm to convert Dendra2 from its “green” state (excitable by 488 nm light) to its “yellow” state (excitable by 561 nm light) as well transferring dark Alexa647 molecules to the fluorescent state. A “blinking buffer” (see methods chapter 8.4 and 8.5) was used to switch Alexa647 back into the dark state. With 561 nm excitation, PALM imaging of Dendra2 could be performed, yielding super-resolution images of CinW2. Excitation at 647 nm was be used for STORM imaging of Alexa647 and results in CinY2 super-resolution images.

In order to determine whether the GFPTrap was binding non-specifically to Dendra2 or only specifically to eGFP, fluorescence correlation spectroscopy (FCS) measurements have been carried out to extract diffusion constants. To this end, three samples using PBS as a buffer have been created: GFPTrap alone, purified eGFP incubated at a ten-fold excess with the GFPTrap, and purified Dendra2 incubated at a ten-fold excess with the GFPTrap.

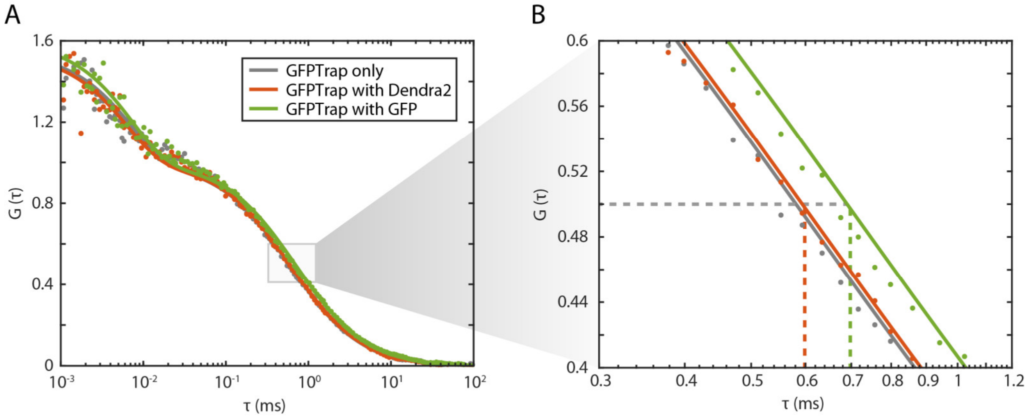


Figure 6.2 FCS measurements of the GFPTrap with eGFP and Dendra2 A) FCS curves of the GFPTrap alone as well as incubated with Dendra2 and eGFP detected using 488 nm excitation light. The curves show the typical triplet-state peak below 10^{-2} milliseconds and the sigmoidal shape of the diffusion process in the millisecond range. B) Zoom-in of the highlighted area in A), displaying the differences in the diffusion time.

The GFPTrap alone serves as a negative control and yielded a diffusion time of $\tau_{dif}^{trap} = 0.60 \pm 0.01\ \text{ms}$, corresponding to a diffusion constant of $D_{trap} = 149 \pm 2\ \mu\text{m}^2/\text{s}$ (Figure 6.2, gray lines) which is in line with reported values for nanobodies of similar size [108]. Incubation with purified Dendra2 did not change the diffusion behavior significantly ($\tau_{dif}^{trap+Dendra2} = 0.61 \pm 0.01\ \text{ms}$, $D_{trap+Dendra2} = 146 \pm 2\ \mu\text{m}^2/\text{s}$, red in Figure 6.2), indicating no binding activity between the GFPTrap and Dendra2. An incubation with eGFP, however, increased the diffusion time to $\tau_{dif}^{trap+eGFP} = 0.71 \pm 0.01\ \text{ms}$ and decreased the diffusion constant to $D_{trap+eGFP} = 125 \pm 2\ \mu\text{m}^2/\text{s}$, confirming a binding activity and the correct function of the GFPTrap (Figure 6.2, green lines).

6.2 Steps towards PALM-STORM: screening, alignment, and imaging routine

After introducing the transformation workflow in the last subchapter, this chapter will explain the steps towards the final two-color SMLM images. In the following chapters and figures, the green channel will always represent Dendra2 coupled to CinW2, while the red channel will represent Alexa647 coupled to CinY2. The CinY2-eGFP-GFPTrap-Alexa647 stack will be abbreviated CinY2-Alexa647.

Screening. After the first transformation and introduction of CinY2-eGFP, a successful cloning could be identified by eGFP fluorescence in the 488 nm channel. Figure 6.3A (top row) shows a distinct fluorescence signal in the girdle band region of the diatom in the 488 nm channel, as expected from the cingulin occurrence inside the organism. The 561 and 647 nm channels only displayed chloroplast autofluorescence. These patterns do not change with UV illumination, except for slight bleaching of the chloroplasts (Figure 6.3A, bottom row).

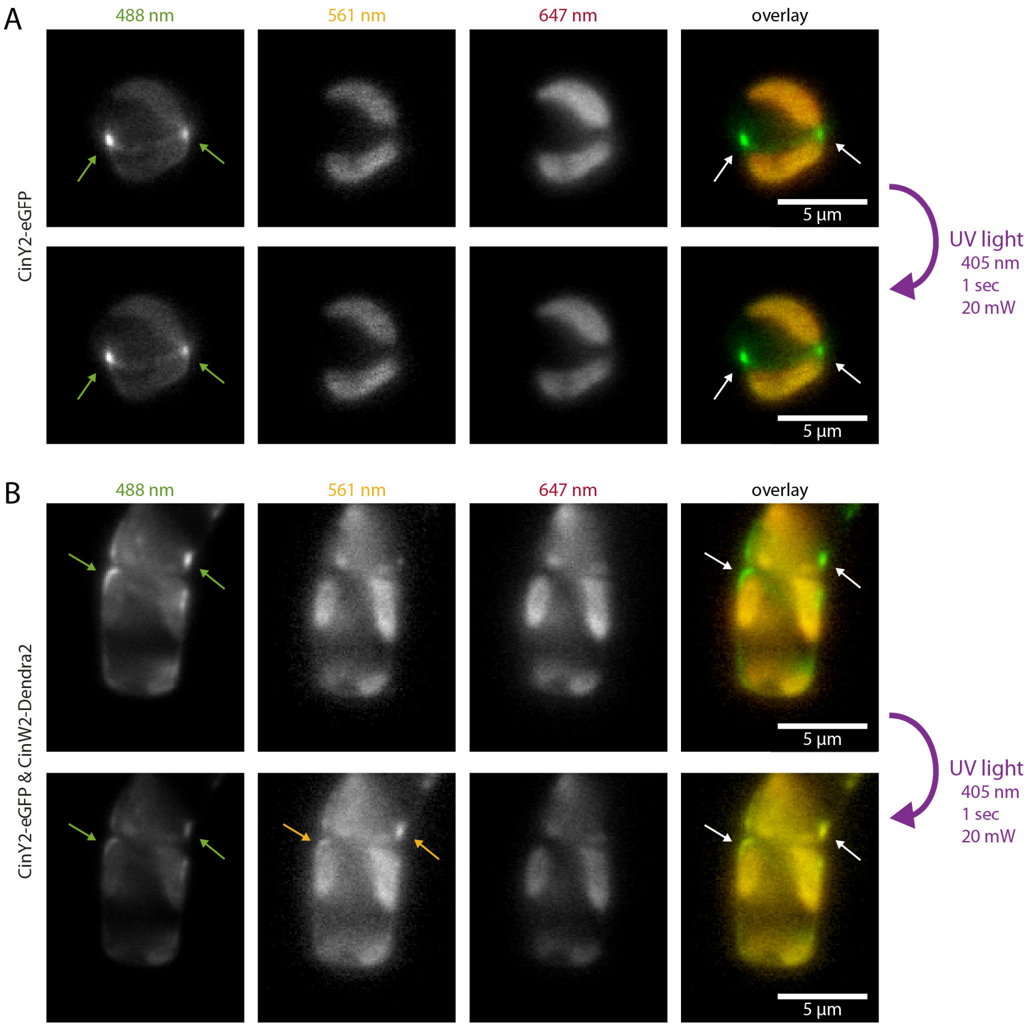


Figure 6.3 Widefield images of live diatoms screened for eGFP and Dendra2 fluorescence The four columns display the three different excitation channels as well as a colored overlay of all three channels. The eGFP and Dendra2 signal is highlighted with arrows. A) In the first screening round, a successful transformation for eGFP is displayed as a distinct fluorescence signal in the green channel. After UV illumination (bottom row), the signal stays unchanged. B) In the second screening round for Dendra2, the successful transformant displays a fluorescent signal in the green channel before UV illumination (top row) as well as a signal in both the green and yellow channel after UV illumination (bottom row).

In the second transformation step, CinW2-Dendra2 was introduced. A positive cloning shows similar fluorescence patterns in the 488 nm channel, where the two Cingulin species tagged with eGFP and (unconverted) Dendra2 will overlap. Therefore, the 561 nm channel has to be used to identify UV-converted Dendra2. Figure 6.3B shows the increase of 561 nm fluorescence signal after UV activation (bottom row). The shift of Dendra2 fluorescence from its green to yellow state is apparent in the overlay projection.

Microring widefield imaging. After lysis, ammonium fluoride treatment, and immunostaining as described in chapter 6.1, the microrings were imaged using 561 nm illumination for CinW2-Dendra2 and 647 nm for CinY2-Alexa647 in TIRF mode. Figure 6.4 displays widefield images of multiple microrings for the two channels as well as a colored overlay projection. The ring-like structure is clearly visible, although with major deformations, as the stabilizing biosilica shell is not present anymore. From the widefield images alone, it is hardly possible to obtain cingulin specific pattern information as the channels show almost identical images. Therefore, super-resolution imaging is necessary to reveal if the apparent co-localization is also present in the nanometer regime.

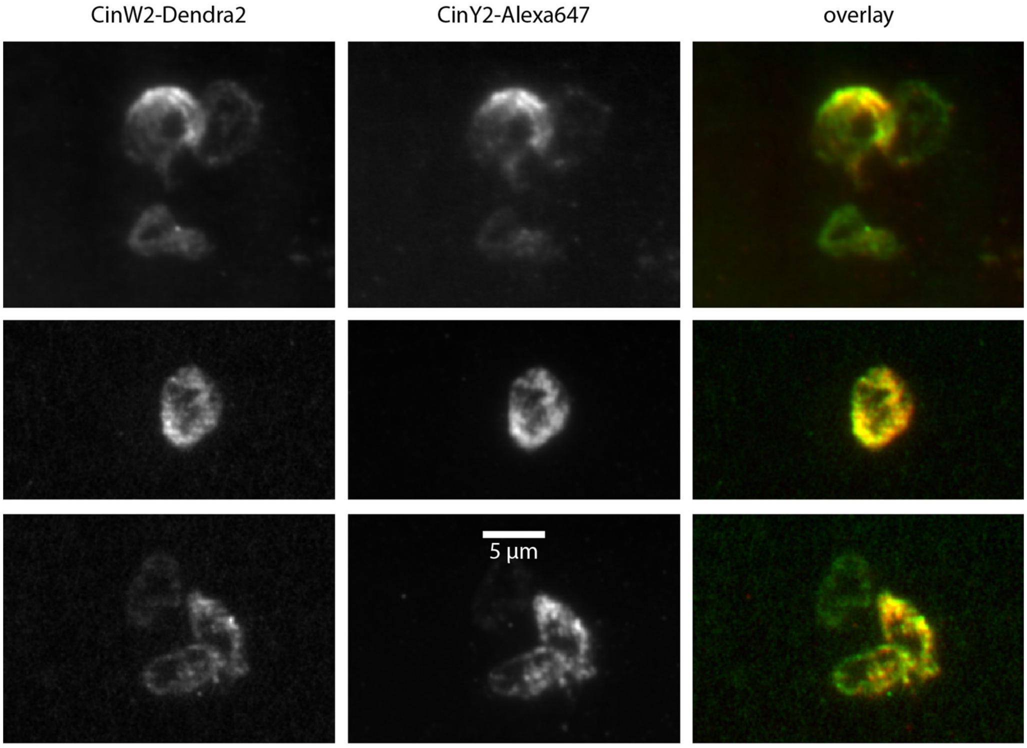


Figure 6.4 Widefield images of CinW2 & CinY2 tagged microrings The three columns represent widefield images of CinW2-Dendra2, CinY2-Alexa647 and an overlay of both channels, respectively.

Microring super-resolution imaging. For super-resolution imaging, the two channels have been recorded consecutively for at least 1000 frames. After spot-finding with ThunderSTORM, a drift-correction was applied as developed for the DNA origami test samples (chapter 3). Figure 6.5 shows overviews of the resulting STORM and PALM microring images.

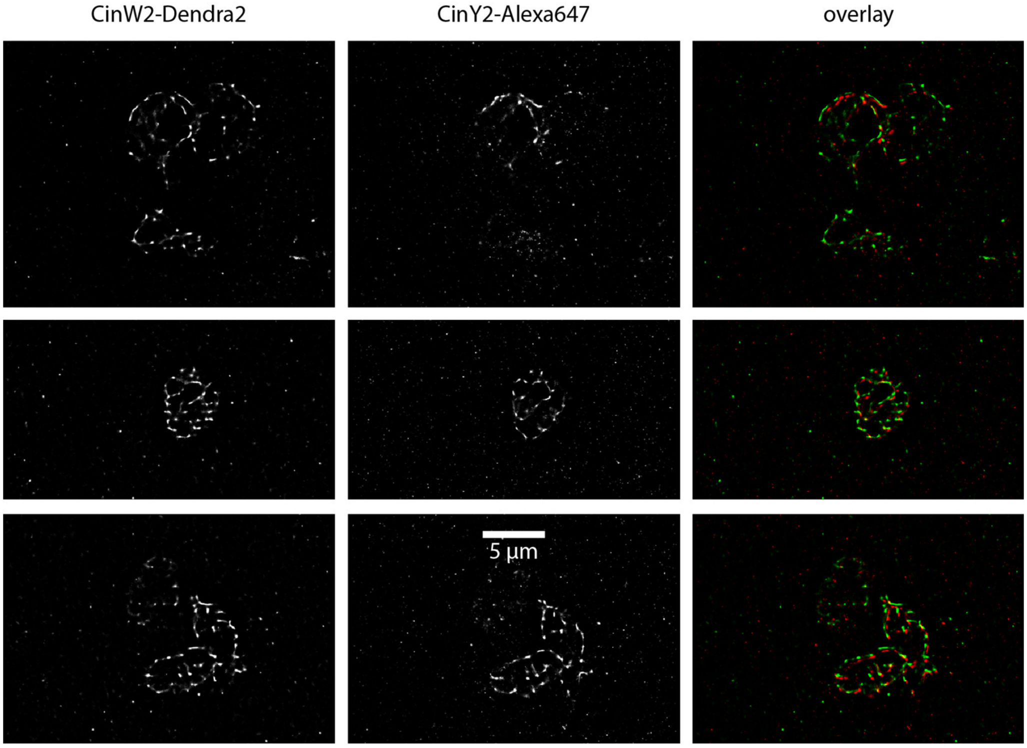


Figure 6.5 Super-resolution images of CinW2 & CinY2 tagged microrings The three columns represent super-resolution images of CinW2-Dendra2, CinY2-Alexa647 and an overlay of both channels, respectively. The same microrings are displayed as in Figure 6.4.

Filamentous structures are clearly visible. The two channels are highly correlated and overlapping, although there is an offset due to drift between the measurements, which will be corrected.

Channel clean-up and alignment. In order to align the two channels and to remove noise (present as randomly distributed localizations stemming from unspecific binding of fluorophores to the surface), the following workflow – as outlined in Figure 6.6 - has been applied:

1. Noise in the dataset was detected and removed in each individual channel as described in chapter 1 with the help of Voronoi tessellation.

2. Both channels were aligned with a MATLAB routine that registers two point clouds using the ICP (iterative closest point) algorithm (called `pcregrigid`) by a translational and rotational transformation.
3. The final filtered and aligned localizations were plotted and overlaid for qualitative visual interpretation and further analyzed towards co-localization studies.

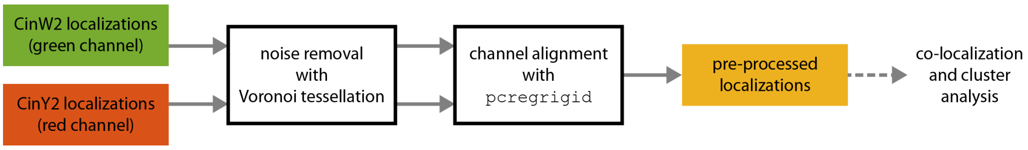


Figure 6.6 Channel clean-up and alignment workflow Flow chart displaying the analysis routine starting from individual super-resolution localization files of two channels resulting in filtered and aligned two-color datasets. These datasets are then ready for further analysis steps after applying noise removal with Voronoi tessellation and a channel alignment. The individual steps are explained in detail in the main text and the corresponding MATLAB algorithm names are written in `typewriter` font.

The results of this workflow are demonstrated in Figure 6.7. Before applying Voronoi tessellation, random localizations (noise) can be seen around the dense cingulin filaments in both channels (Figure 6.7A). These localizations were identified and removed (gray localizations in Figure 6.7B). This filtered dataset was overlaid in Figure 6.7C (top), showing an offset that has been corrected in Figure 6.7C (bottom).

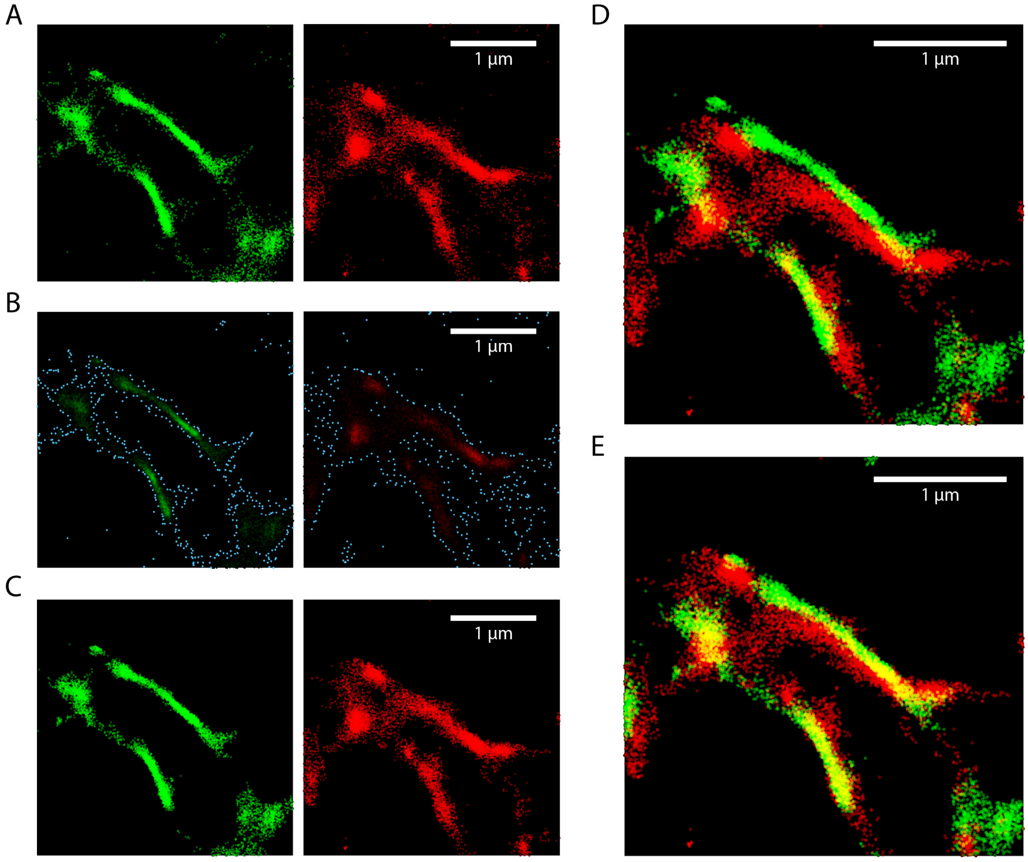


Figure 6.7 Super-resolution scatter plot of microrings before and after filtering and alignment A) Scatter plot of all localizations in the green and red channel. B) After Voronoi tessellation, noise can be identified (plotted in blue). C) Resulting scatter plot without noise. D) Overlay of both channels shows an offset due to drift between the channels. E) After channel alignment, the two channels co-localize.

Due to chromatic aberrations there is still a remaining offset between the channels which can be quantified using fluorescent beads. For the present microscope system and imaging conditions, the chromatic aberrations have been found to be between 0 and 100 nm, depending on the distance from the image center. As all microrings have been centered before imaging and display a dimension of below 10 µm, these chromatic aberration will account for an error below 20 nm. A detailed explanation of the quantification routine can be found in the appendix chapter 9.2.4.

Qualitative Cingulin distribution. In order to assess the filamentous structures in more detail, zoom-ins for several regions are shown in Figure 6.8 in a scatter plot representation. Dense filaments are visible in both channels with a high degree of co-localization.

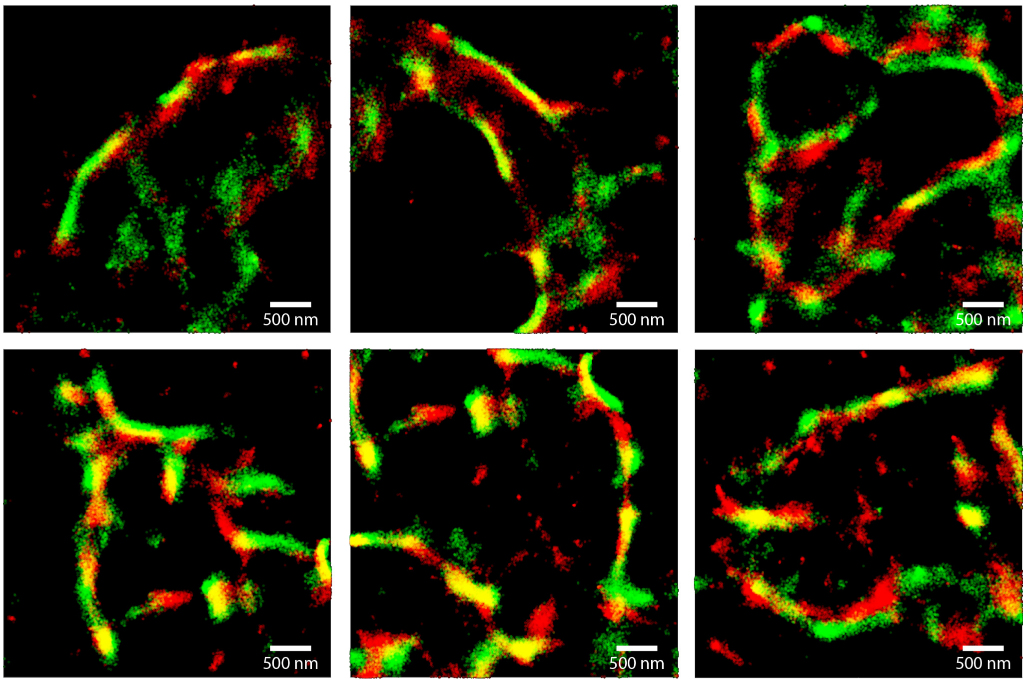


Figure 6.8 Super-resolution scatter plot details of microrings The scatter plots reveal filamentous cingulin structures. CinW2-Dendra2 localizations are displayed in green and CinY2-Alexa647 in red.

Prior to quantitatively evaluating these patterns and the co-localization of the cingulins, it is important to note that while CinY2 has been localized using Alexa647 which yields a higher brightness and therefore a higher localization precision than Dendra2, it also features a longer linker between the cingulin and Alexa647 (consisting of eGFP and the GFPTrap) which decreases the final localization precision of CinY2. In order to asses if these factors might influence the final localization pattern, it is necessary to take a closer look at the localization parameters.

Localization parameters. As expected, the localization precision of Alexa647 with 16.1 nm is better than that of Dendra2 with 20.5 nm (Figure 6.9A). On the other hand, the FRC yields very similar resolution estimates for both channels at around 57 nm (Figure 6.9B). This suggests that the better localization precision of Alexa647 “compensates” the longer linker resulting in a similar final image resolution as defined by the FRC. This means that the two channels can be compared quantitatively without adding any correction terms. For example, an apparently thicker filament in the super-resolution image could be an effect of a decreased resolution in that specific channel, which can be ruled out in the aforementioned scenario. Finally, to ensure a proper alignment of both channels the nearest neighbor in the opposite channel has been calculated for each localization and is plotted in Figure 6.9C. For two properly aligned and co-localizing channels, a peak at low distances is expected. Indeed, the given histograms show a peak at around 5 nm for both channels which is well below the localization precision, confirming both dense localizations and well aligned channels.

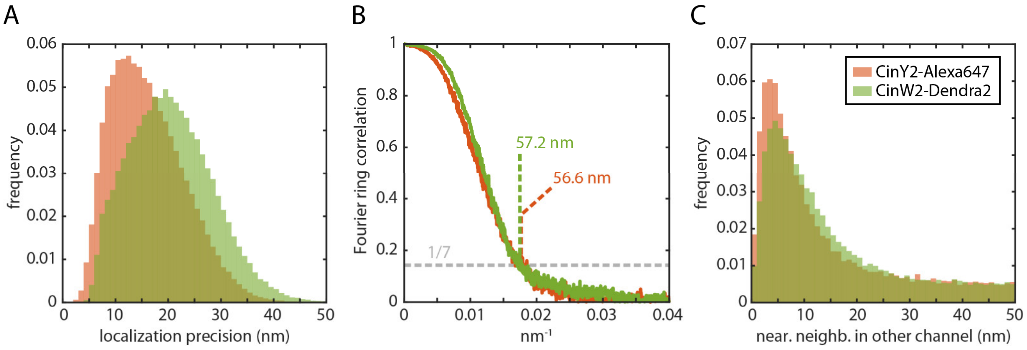


Figure 6.9 Localization precision and nearest neighbor for both channels A) Localization precision distribution for the red and green channel resulting in a mean of 16.1 and 20.5 nm, respectively. B) Fourier ring correlation resulting in resolution estimates of 56.6 nm (red channel) and 57.2 nm (green channel). C) Distribution for the nearest neighbor distance in opposite channels peaking at around 5 nm for both channels, indicating very good channel alignment.

6.3 Co-localization studies: quantification, clustering, and correlations

After a successful channel filtering and alignment, co-localization studies can be performed to elucidate differences in protein patterns between the two cingulin species. The schematic in Figure 6.10 displays the different analysis routines and MATLAB algorithms that were applied. First, the localizations were analyzed qualitatively and quantitatively by applying a coordinate-based co-localization (CBC) algorithm as described in chapter 2.5. Subsequently, by clustering the data with DBSCAN (chapter 2.4), access to cluster-specific information like shape, abundancy and size was gained that can be compared between the two channels resp. between the two proteins.

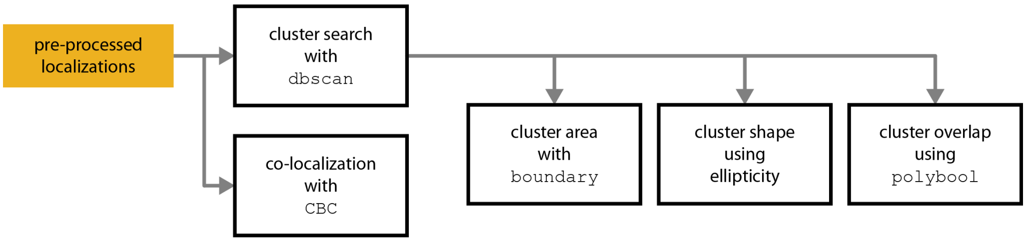


Figure 6.10 Co-localization and cluster analysis workflow Flow chart displaying the analysis routine starting from a pre-processed two-color dataset with different options of cluster analysis and/or co-localization studies. The individual steps are explained in detail in the main text and the corresponding MATLAB algorithm names are written in `typewriter` font.

CBC input parameter optimization. As described in [67], and introduced in chapter 2.5, CBC features two free parameters: the binning Δr and the maximum search radius R_{max}. All localizations are evaluated by the number of neighboring localizations within search radii r from zero to R_{max} using the step size Δr. According to this number, a normalized co-localization value C is calculated that ranges from -1 (anti-correlation) to +1 (perfect co-localization). By visualizing the co-localization value for different combinations of Δr and R_{max}, optimized parameter values can be found visually applying the following procedure (Figure 6.11):

- A high Δr of 50 nm (Figure 6.11A, bottom) causes a loss in structural details since not enough distances are sampled. Additionally low correlation values disappear.
- A low Δr of 2 nm (Figure 6.11A, top) which is significantly below the localization precision of 16.1 nm will blur and smoothen out hard steps in the co-localization value and thereby decrease the local contrast.
- A high R_{max} of 1000 nm (Figure 6.11C, bottom) will introduce false positive co-localization values as the radius is bigger than single clusters or structures in the image.
- A low R_{max} of 50 nm (Figure 6.11C, top) causes a loss in correlation as the area that is sampled is too small resulting in parts of the structure having no correlation information at all (i.e. a C value of zero).

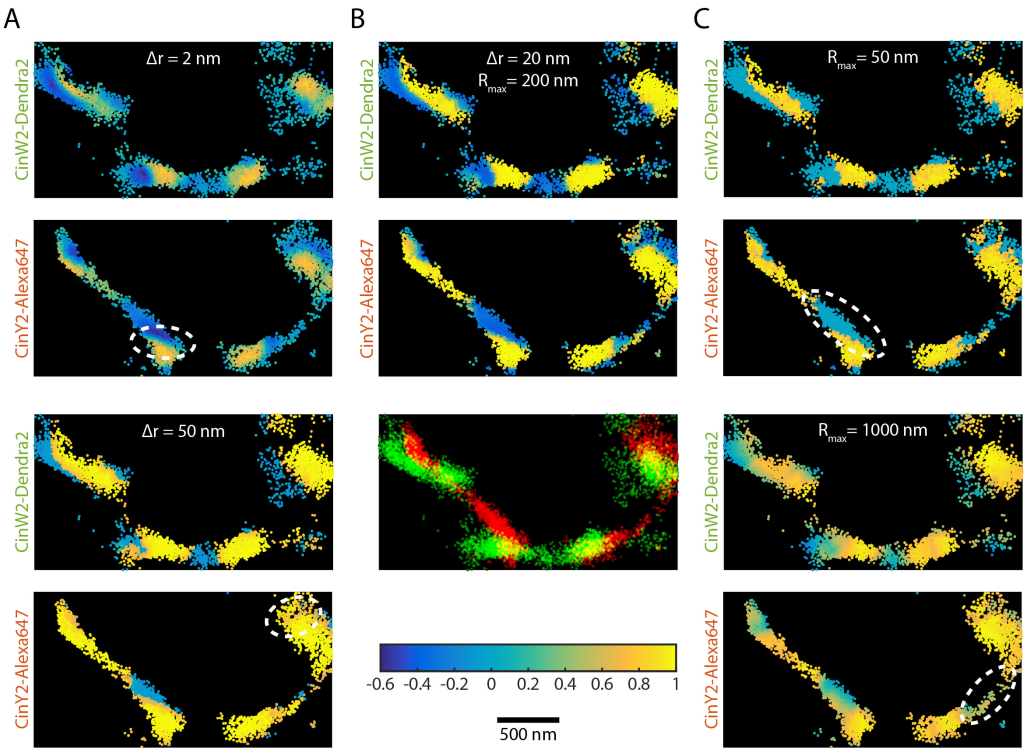


Figure 6.11 Influence of CBC parameter The influence of the two CBC parameter Δr and R_{max} is displayed using scatter plots of the same region of interest, colored by the co-localization value. If not indicated otherwise, the default parameter values are Δr = 20 nm and R_{max} = 200 nm. Columns A) and C) display changes in r and R_{max}, respectively. The areas of interest which are discussed in the main text are highlighted with white circles. Column B) shows the scatter plot for optimized CBC parameters as well as the two-color scatter plot as a reference.

By probing these different parameter combinations, an optimal value pair of are Δr = 20 nm and R_{max} = 200 nm has been found (Figure 6.11B), which is in line with the "rule of thumb" given in the original publication, stating that Δr should be around the PALM / STORM resolution while R_{max} is chosen in the range of the expected structure / cluster size. These parameters result in co-localization values of high contrast that resolve fine structural details while showing no apparent co-localization artifacts.

Qualitative CBC analysis. The following scenarios have been observed by analyzing the co-localization values qualitatively in scatter plots of multiple microrings (Figure 6.12):

- The most apparent feature is the high co-localization between both channels resp. proteins (Figure 6.12A). The cingulin filaments are similar in shape and overlap in large parts with only fine structures being different.

- Slight differences could be observed from microring to microring regarding the filament elongation and thickness, but no general trend was detectable. In Figure 6.12B-left, the CinY2-Alexa647 filament appears slightly thinner. Figure 6.12B-right shows a CinW2-Dendra2 filament that is more defined while CinY2-Alexa647 is distributed in a broader fashion.
- CinW2-Dendra2 filaments display a higher continuity and therefore longer elongation compared to CinY2-Alexa647 which often exhibits smaller patches and clusters (Figure 6.12C). In the right panel of Figure 6.12C, CinY2-Alexa647 forms almost circular clusters instead of a filament.

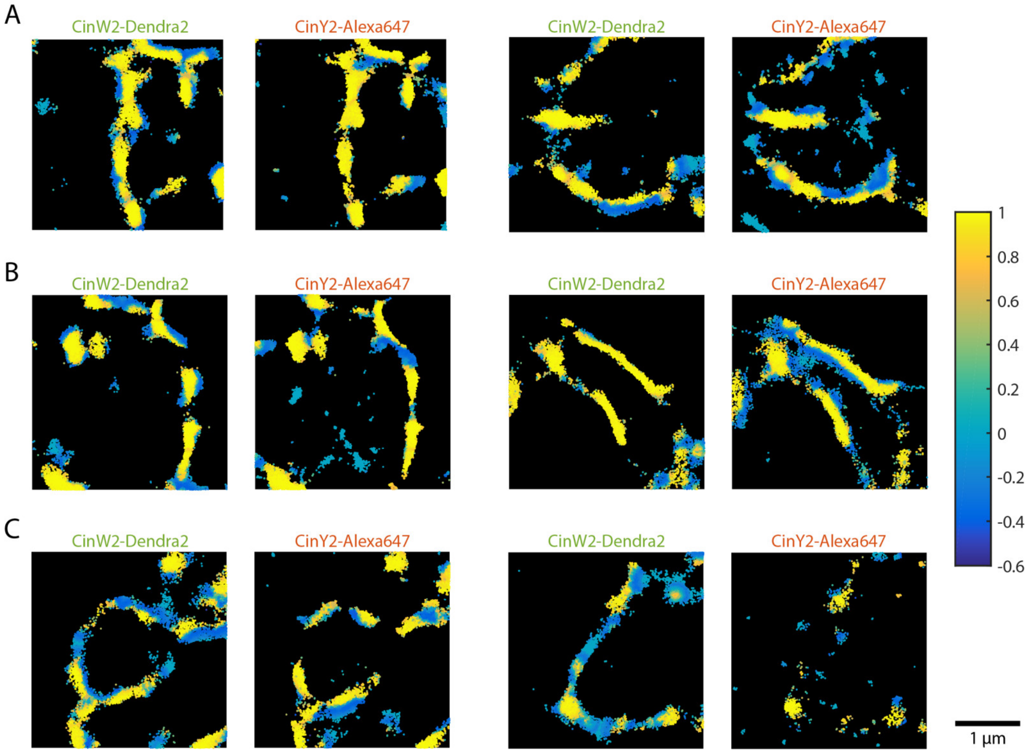


Figure 6.12 Co-localization scatter plots Scatter plots of microrings for CinW2-Dendra2 and CinY2-Alexa647, colored by the co-localization value are shown. A) Highlighting of thr high co-localization degree between the channels. B) Display of the differences in filament elongation and thickness. C) Depiction of the “patchy character” of CinY2-Alexa647 localizations.

In order to evaluate whether a certain scenario is dominant and to establish a connection between the observed patterns and the functional role of the cingulins, a more quantitative analysis is necessary.

Quantitative CBC analysis. In the first place, the co-localization value from multiple images was pooled to form a histogram (Figure 6.13A). The majority of localizations in both channels displayed a high co-localization value as the scatter plots (Figure 6.12A) already suggested. A second population at co-localization values around -0.3 indicates the presence of clusters unique to a single channel (i.e. areas where only a single cingulin is present). By plotting the difference between the co-localization values from both channels, it is possible to highlight disparities in the distributions efficiently (Figure 6.13B). While CinY2-Alexa647 localizations are more dominant at co-localization values above 0.8, CinW2-Dendra2 features more anti-correlated localizations.

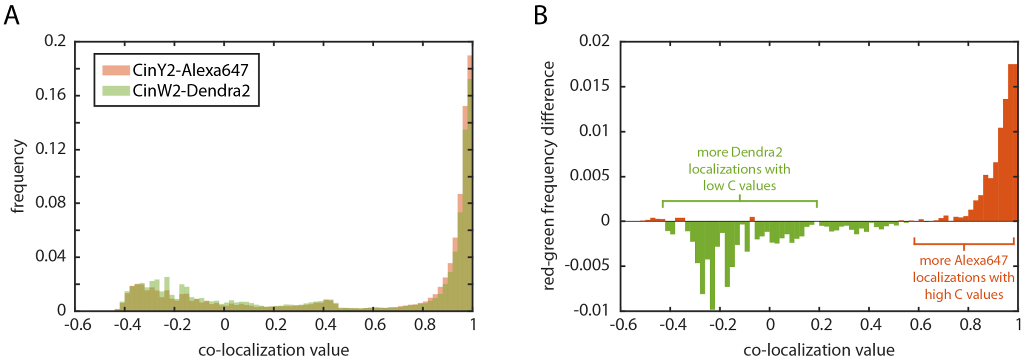


Figure 6.13 Co-localization histograms A) Normalized histogram of the co-localization value C for both channels. B) The difference between the histograms from A) displays co-localization value preferences in different regions.

This supports a trend that could already be observed in Figure 6.12C, namely that – on average – CinY2-Alexa647 forms smaller and spottier clusters that co-localize with the more elongated and longer CinW2-Dendra2 filaments. In order to analyze clustering and co-localization in more detail, individual clusters from the dataset were extracted.

Cluster search and shape analysis. Cluster search has been applied using DBSCAN as introduced in chapter 2.4. It is best suited for clusters of unknown geometry as is the case for the cingulin localizations. By using a distance threshold ε of 60 nm, which corresponds to the image resolution obtained by FRC, the algorithm yielded meaningful results (Figure 6.14) that effectively split the dataset into multiple clusters for each of the two channels.

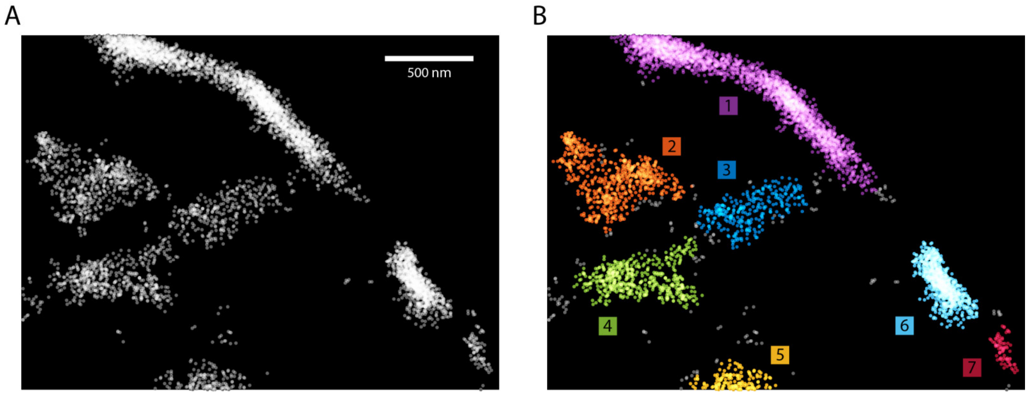


Figure 6.14 DBSCAN for cingulin cluster identification A) Exemplary scatter plot of a cingulin filament showing distinct areas of dense localizations. B) After applying DBSCAN, each localization is assigned to an individual cluster and can be further processed.

In order to compare the set of clusters between both channels it was necessary to extract the area of the clusters and parameter to define the cluster shape. Normally, the area of a set of points is defined by the convex hull. As Figure 6.15A shows, this convex hull greatly overestimates the area occupied by the point cloud if its shape has concave features. As this is often the case for the presented cingulin filaments, a different approach had to be used. So called alpha shapes generalize the concept of convex hulls to a more intuitive notion of "shape" by creating a tight boundary around the point cloud [109]. By applying this algorithm (using the MATLAB function `boundary` with the default shrinkage factor of 0.5) to the clusters their boundary could be effectively approximated, as shown in Figure 6.15B, and their area could be accurately calculated.

As the cingulins preferentially form longer filaments, elongated cluster shapes were expected. In order to quantify this "elongation" and compare the two studied proteins it is necessary to find a parameter that reflects this shape. By fitting the whole point cloud of a single cluster to a two-dimensional Gaussian distribution it was possible to define an ellipticity by the ratio of the eigenvalues from the covariance matrix of this Gaussian. As Figure 6.15C shows, the point cloud cannot be represented accurately with a single 2D Gaussian. Yet, it gives a good approximation of its elongation and ellipticity.

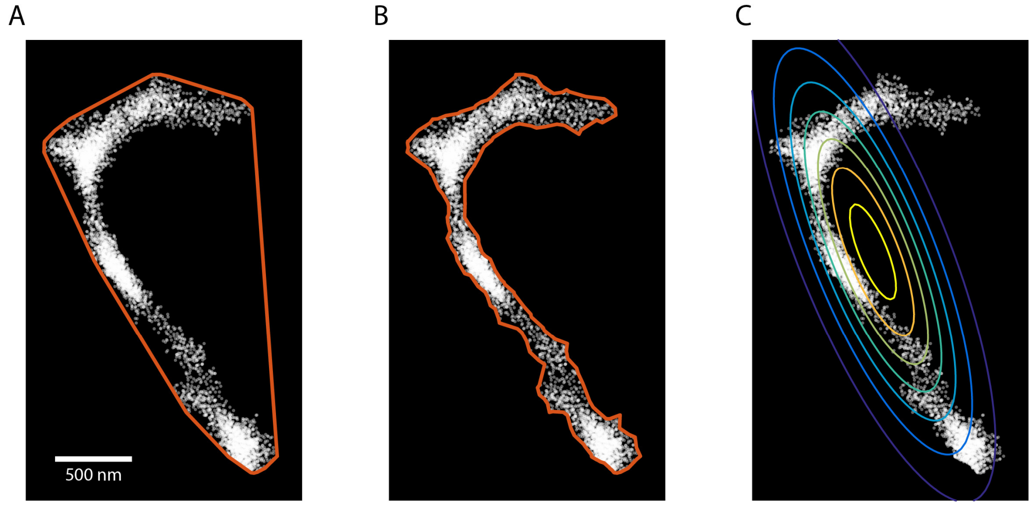


Figure 6.15 Defining the cluster area and cluster ellipticity A) Scatter plot of a single filamentous structure. A simple area calculation using the convex hull (MATLAB function `convhull`) will overestimate the occupied area massively. B) Using alpha shapes (MATLAB function `boundary`), a tight hull can be defined that resembles the overall structure. C) The ellipticity of the structure can be estimated by fitting a single elliptical Gaussian to the point cloud.

The clusters from measurements of seven microrings have been pooled (182 and 286 clusters total in the green and red channel, respectively) and analyzed with regard to size and shape. Figure 6.16 displays histograms of the calculated cluster area and ellipticity. See Table 4 at the end of this chapter for an overview of all the obtained values. The mean cluster area for CinW2-Dendra2 with $0.16 \pm 0.02\ \mu m^2$ (standard error) is twice the size of CinY2-Alexa647 ($0.08 \pm 0.01\ \mu m^2$). CinY2-Alexa647 displays 50% more clusters in total. The mean ellipticity of CinW2-Dendra2 with 8.2 is 57% higher than for CinY2-Alexa647 (5.2).

In summary it can be concluded that CinW2-Dendra2 forms larger but fewer clusters that are more elongated (or less spherical) than CinY2-Alexa647. This supports the CBC values that already indicated that CinW2 is a longer filament.

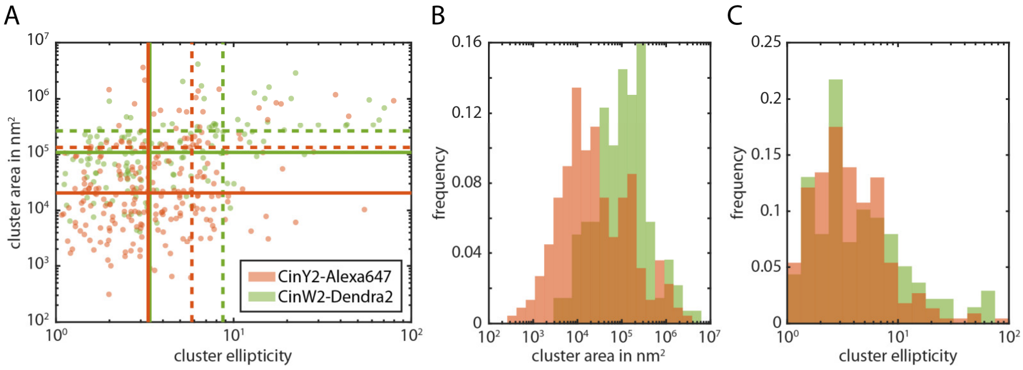


Figure 6.16 Area and shape statistics of the detected clusters Statistics for the clusters found with DBSCAN in both the green (CinW2-Dendra2) and red (CinY2-Alexa647) channel. A) Scatter plot of the cluster ellipticity and area. Solid and dashed lines represent the median and mean, respectively. B) Histogram of the cluster area. C) Histogram of the cluster ellipticity.

Cluster overlap analysis. A final analysis has been carried out with regard to co-localization of the detected clusters as an alternative approach to CBC to confirm the obtained results. The amount of co-localization of multiple shapes is given by the size of the overlapping region between them. Mathematically speaking, it is the Boolean intersection of polygons defined by the boundary of clusters (using the MATLAB function `boundary`). Figure 6.17A highlights an exemplary region of interest featuring strongly overlapping clusters of cingulin localizations. In panel B, the cluster overlap, as calculated using the MATLAB function `polybool`, is highlighted in yellow.

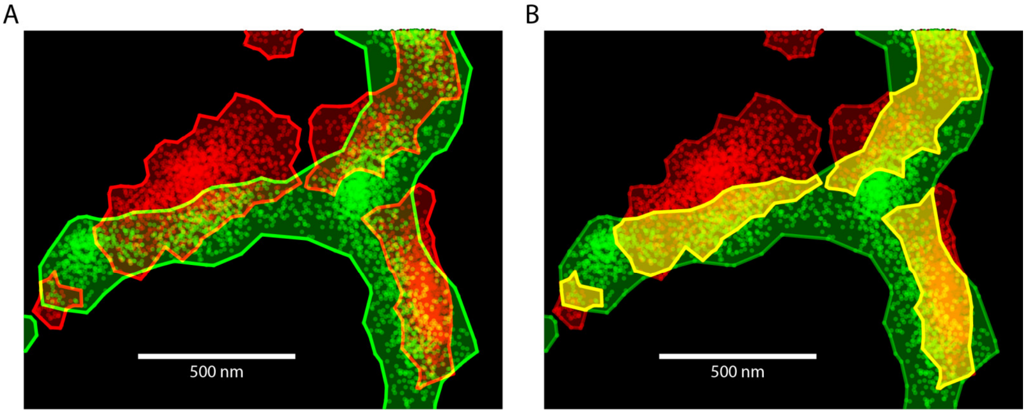


Figure 6.17 From two-color clusters to cluster overlap A) Scatter plot showing CinW2-Dendra2 localizations in green and CinY2-Alexa647 localizations in red. Detected clusters are highlighted. B) The overlap region (MATLAB function `polybool`) of the clusters is highlighted in yellow.

This procedure resulted in a total of 183 overlapping clusters, as summarized in Table 4. Furthermore, the total overlap area adds up to 9 µm^2, yielding a relative overlap area of 32% and 41% compared to the total cluster area for CinW2-Dendra2 and CinY2-Alexa647, respectively. In other words, the CinY2 channel features 28% more overlap with the other channel. This is in line with the CBC results, where CinY2-Alexa647 featured a higher fraction of localizations with a co-localization value above 0.8.

Table 4 Cluster and overlap statistics Statistics for the clusters found in each channel, as well as the overlapping clusters created by pooling data from seven observed microrings.

	CinW2-Dendra2	CinY2-Alexa647	Overlap clusters
Number of clusters	182	286	183
Total cluster area (µm^2)	28.6	22.2	9.0
Mean cluster area (µm^2)	0.16	0.08	0.05
Mean ellipticity	8.2 (800*200 nm)	5.2 (450*170 nm)	
Overlapping region (%)	32	41	

All the values shown in the table have been tested with different DBSCAN settings. While a less stringent cluster search naturally results in bigger clusters, the ratio between the two channels remains intact (i.e. for a distance threshold of 100 instead of 60 nm, the mean cluster area is 0.26 and 0.13 µm^2 with a mean ellipticity of 8.7 and 5.8 for the CinW2 and CinY2 channel, respectively).

6.4 Conclusion

After single-color PALM in diatoms had been established as presented in the previous chapter, it could now be shown that the extension to two-color SMLM yields new insights into spatial correlations of proteins (i.e. cingulins) on the nanoscale.

Due to the limited number of PCFPs that are functional when encased in biosilica and the chloroplast autofluorescence, a combined PALM-STORM approach had to be chosen. CinW2-Dendra2 as a PALM probe and CinY2-eGFP-GFPTrap-Alexa647 as a STORM probe could be localized with an average precision of 21 nm and 16 nm, respectively. Multiple quality checks have been carried out to ensure a high image resolution (by using FRC to estimate the resolution), good signal to noise ratio (by removing localization outliers with Voronoi tessellation), and proper channel alignment (by comparing nearest neighbors in opposite channels).

Besides the coordinate-based co-localization (CBC), multiple analysis routines have been developed in regards to clustering (including the cluster search, cluster area estimates, and cluster shape) and co-localization (by calculating the cluster overlap). By combining these techniques, the recorded SMLM images could be analyzed in detail and revealed the following features (see Figure 6.18 for a visual overview): Both the CinW2 and the CinY2 channel are highly co-localized, which has been observed both in the CBC analysis (featuring high C values) and the cluster overlap (displaying a high degree of overlap). By comparing cluster abundancy and area it could be shown that CinW2-Dendra2 forms larger but fewer clusters, while CinY2-Alexa647 forms more clusters that are smaller. This already indicates a difference in the spatial patterns and ultimately the function of these two proteins. The cluster shape analysis confirmed this conjecture by revealing that CinW2-Dendra2 clusters are more elongated and less circular than CinY2-Alexa647 clusters. As CinY2-Alexa647 overlaps 28% more with the opposite channel, one can hypothesize that it acts as a local supporting structure in the microrings, perhaps at intersections, due to its smaller cluster size and patchy appearance. On the other hand CinW2-Dendra2 displays more elongated structures that form majorly continuous filaments and therefore may act as a stabilizing and connecting element.

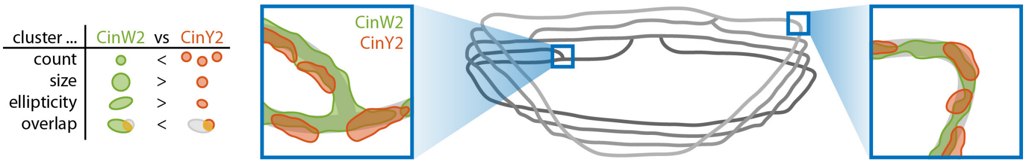


Figure 6.18 Cingulin cluster parameter overview and model hypothesis The table on the left side gives a comparative visual overview of the main cingulin cluster parameters for CinW2-Dendra2 and CinY2-Alexa647. On the right, simplified model distributions are displayed, visualizing and abstracting the observed cingulin patterns.

7 OUTLOOK

The cingulin studies presented in chapter 6 showed that a combination of STORM and PALM can be used to image two cingulins simultaneously. This was achieved by creation of a CinW2-Dendra2 fusion protein to generate a PALM probe and by immunostaining of a CinY2-eGFP construct with a GFPTrap-Alexa647 nanobody to generate a STORM probe. In order to extend this protocol towards three color imaging, it would be necessary to use the eGFP channel (488 nm excitation) for another STORM probe like Alexa488. This could be achieved by replacing the eGFP fusion protein as the antigen for the nanobody by a simple small tag like His-tag, HaloTag or SNAP-tag. This would have the added benefit that the final linker is much smaller (e.g. CinY2-SNAP-Alexa647 instead of CinY2-eGFP-GFPTrap-Alexa647) and thus, the localization precision would be higher. Figure 7.1 shows an exemplary labeling scheme. However, this approach would necessitate a “blind” screening for positive clones which would require to massively increase the number of screened clones. In addition to the described construct containing CinY2-eGFP and CinW2-Dendra2, constructs with other cingulin combinations could be created. By keeping one cingulin constant, it could act as a reference cingulin channel to be compared with one of the other five cingulins.

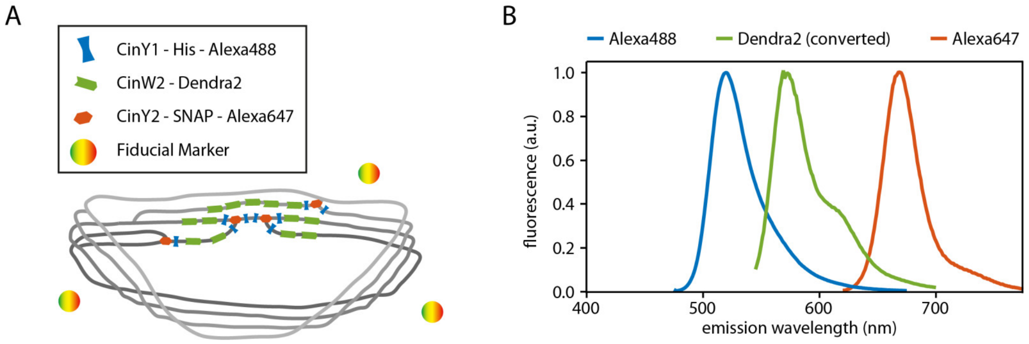


Figure 7.1 Proposal for a three-color cingulin construct A) Exemplary labeling scheme to image three labeled cingulins at the same time. Fiducials would allow easy channel alignment as well as subsequent correlative electron microscopy. B) Emission spectra of the proposed fluorescent probes Alexa488, Dendra2 (after photo conversion), and Alexa647. The spectral separation enables three-color imaging (using 488, 561, and 647 nm excitation light).

In the described approach, only the labeled components of the microrings appear in the fluorescence images. A reference image of the complete organic filament would be desirable to prevent the misinterpretation of gaps in the fluorescence image as a gaps in the microring. This could be achieved by combining SMLM with other imaging techniques like atomic force microscopy (AFM) or electron microscopy. In AFM studies, the microring fibers could be fully visualized with a height of around 6 nm [17]. This characteristic height profile could be used as a mask to visualize the complete microrings in these correlative SMLM–AFM measurements. Certainly, suitable fiducials to correlate the two channels are necessary and alignment is very critical [110]. Similarly, TEM could be applied to image the microrings [111]. Alternatively, not only the microrings, but the complete biosilica cell wall could act as a reference channel that can be imaged with SEM [112]. This requires to omit the biosilica dissolution step and would make it possible to image cingulins "*in situ*" in their native environment in correlation to the intact biosilica cell wall topography. For these studies, obtaining 3D localization information is advantageous, if not crucial. This could be achieved by introducing a cylindrical lens into the emission pathway: without major modifications to the imaging system, 3D information down to an axial resolution of about 50 nm can be obtained [41]. It remains to be seen if the current signal to noise ratio of the SMLM images is sufficient to extend the imaging procedure into the third dimension. Naturally, these correlative imaging techniques could be extended towards studies of any fluorescently tagged and even biosilica embedded component within *T. pseudonana*.

It is highly desirable to take this "*in situ*" approach one step further and establish *in vivo* SMLM in live diatoms. With the approach pursued in this work for silaffins (chapter 5) and cingulins (chapter 6), obtaining *in vivo* SMLM images was hampered due to the chloroplasts autofluorescence and scattering of light inside the diatom. Mainly, this difficulty correlates with the weak fluorescence of the fusion protein. Very recent technical developments might indicate ways to circumvent this shortcoming. Generally, two strategies could be followed: Either the signal or expression level of the fluorescent probe has to be increased to be sufficiently stronger than the chloroplast fluorescent background, or the background has to be reduced. For the first strategy (increasing the signal of the fluorescent probe), a recently established genetic transformation protocol for *T. pseudonana* based on CRISPR-Cas [26] might allow a more stable, more specific and more efficient expression of the tagged protein. Alternatively, the use of novel PCFPs featuring increased brightness for PALM or the generally much brighter fluorescent dyes for STORM might prove helpful. For example, structured illumination microscopy studies showed that bright biarsenic Cyanine fluorophores (AsCy3 and AsCy3e) can be used for silica embedded fusion proteins [113]. Due to their blinking nature these fluorophores are assumed to be suited for SMLM as well. For the second strategy (reducing the background caused by the chloroplast autofluorescence), recently developed super-resolution techniques based on fluctuation analysis (SOFI) [114] are promising approaches for a suppression of the background signal. Preliminary SOFI experiments on tpSil3-Dendra2 *in vivo* have been carried out and the results are briefly presented in the appendix chapter 9.2.3.

In order to learn more about biomineralization in *T. pseudonana* and diatoms in general, it is necessary to extend SMLM studies also to different cell compartments and targets. For example, biomineralization could be studied at its “genesis” in silica deposition vesicles (SDVs) using membrane labeling. It has been shown that commonly used membrane probes can be used for SMLM [115] and pose an interesting opportunity to label diatom-specific compartments like the SDV. The dynamic nature of these compartments during the cell cycle would make time-lapse imaging under super-resolution conditions very intriguing. In comparison to SMLM, stimulated emission depletion (STED) microscopy offers video-rate acquisition speed [28]. Although, high acquisition speed comes at the cost of spatial resolution, this significantly increased time-resolution at super-resolution conditions would allow visualization of highly dynamic processes. First steps towards time-lapse fluorescence imaging of proteins in *T. pseudonana* and their quantitative evaluation have been taken in a satellite project of this and are shown in the appendix chapter 9.2.1.

In summary, this doctoral work successfully implemented SMLM for protein targets embedded inside diatom biosilica for the first time. The developed protocols were applied to study cingulins, as part of an insoluble organic matrix, to learn more about their involvement in biomineralization in *T. pseudonana*. In co-localization studies, CinW2 and CinY2 localizations and patterns have been compared extensively and a model of their functions was developed. Additionally, not only the established protocols and imaging strategies in this doctoral work can be generalized and applied to other biological questions, but also the developed software analysis tools. These tools reach from the data exploration GUI SiMoNa (chapter 2.7), over PairRice – a novel cluster distance algorithm that has been applied to DNA origami nanostructures (chapter 3) – to various cluster analysis tools developed for co-localization studies on cingulins (chapter 6). All these analysis tools are highly flexible and thus not limited to diatom imaging. Therefore, they can be easily adapted and extended for future SMLM studies.

8 MATERIALS & METHODS

8.1 Microscope specifications

The microscope used for all fluorescence microscopic experiments was a Nikon N-STORM bundled with an additional spinning disc unit for confocal microscopy. Since this is a commercial bundle, this section will not go into the details of building up a super-resolution microscope, but will only list all relevant components with its key properties.

The main body of the microscope is a Nikon Eclipse Ti-E inverted microscope system that can be controlled with the Nikon NIS-Elements software.

Excitation sources. For laser illumination, the MLC (monolithic laser combiner) 400 by Keysight, a laser box that threads four diode lasers into one fiber, is available. The four lasers feature the following intensities at the tip of the fiber: 405 nm (20 mW), 488 nm (80 mW), 561 nm (80 mW), 647 nm (125 mW). A TIRF angle actuator allows offsetting the laser path to illuminate the sample at an angle and in total internal reflection, allowing TIRF microscopy. Additionally, a white light lamp is available for bright-field microscopy.

Stage and Focus. The microscope body is equipped with a Nikon motorized x-y-stage (0.1 µm step size, 0.5 µm resolution) as well as a Nano-Z100 z-piezo from Mad City Labs (100 µm range, 0.2 nm resolution). Multiple stage holders are available for 1 by 3 inch microscope slides, petri dishes and multi-well plates. An incubator by Okolab encases the microscope body and stabilizes temperature fluctuations and thereby reduces drift. Additionally it shields the sample from environmental light. The Nikon Eclipse Ti-E features a focus system called “Perfect Focus” that corrects z-drift in real-time and keeps the focus at a constant position over long measurements and stage movements.

Objectives. The following objectives are available for imaging:

- 20x Air: CFI Plan Apochromat VC, NA 0.75, WD 1mm
- 40x Air: CFI Plan Apochromat Lambda, NA 0.95, WD 0.16-0.25mm
- 60x Water: CFI Plan Apochromat IR, NA 1.27, WD 0.15-0.19mm
- 60x Water: CFI Plan Apochromat VC, NA 1.20, WD 0.27mm
- 100x Oil: CFI TIRF Apochromat, NA 1.49, WD 0.12mm

Additionally, a 1.5x tube lines can be inserted into the emission path that increases the magnification by 50%. If not noted otherwise, the 100x objective and the tube lens are used for all measurements in this work.

Widefield emission pathway. One arm of the emission pathway is used for widefield microscopy, including SMLM. An Andor Ixon Ultra 897 EMCCD camera is used for detection (56 fps at 512x512 pixel, pixel size 16 µm). Multiple filter cubes are available that consist of an excitation and emission filter, as well as a dichroic mirror (all by AHF, Chroma):

- 405 nm imaging:
    - F39-390 (laser bandpass 370-410 nm)
    - F48-403 (dichroic longpass 420 nm)
    - F47-460 (emission bandpass 435-485 nm)
- 488 nm imaging:
    - F37-473 (laser bandpass 457-493 nm)
    - F48-486 (dichroic longpass 502 nm)
    - F37-521 (emission bandpass 502.5-547.5 nm)
- 561 nm imaging:
    - F49-555 (laser bandpass 542-568 nm)
    - F48-558 (dichroic longpass 565 nm)
    - F37-609 (emission bandpass 582-636 nm)
- 647 nm imaging:
    - F39-628 (laser bandpass 608-648 nm)
    - F48-643 (dichroic longpass 657 nm)
    - F47-700 (emission bandpass 662.5-737.5 nm)

- N-STORM Filter (allows all excitations but only red emission)
    - no laser bandpass
    - F48-643 (dichroic longpass 657 nm)
    - F47-705 (emission bandpass 669-741 nm)
- Quad band Filter (allows all excitations and all emission)
    - ZET 405/488/561/647 x (quad band laser clean-up filter)
    - ZT 405/488/561/647 rpc (quad band dichroic)
    - ZET 405/488/561/647 m (quad band laser rejection filter)

A cylindrical lens can be inserted into this pathway to allow 3D-SMLM[41].

Confocal emission pathway (spinning disk). The second arm of the microscope body is equipped with a Yokogawa CSU-X1 confocal scanner unit that allows spinning disk confocal microscopy (up to 5000 rpm). Here, an Andor Ixon Ultra 888 is used that features a higher pixel count at a slower acquisition rate than the 897 version (26 fps at 1024x1024 pixel, pixel size 13 µm). The CSU-X1 has an incorporated dichroic mirror to split excitation from emission light that is used for all wavelengths. A filter wheel offers the following imaging options:

- 405 nm: single band filter FF02-447/60-25
- 488 nm: single band filter FF01-525/30-25
- 561 nm: single band filter FF01-607/36-25
- 647 nm: single band filter FF01-685/40-25
- 405 & 488 nm: dual band filter 433/530 HC
- 488 & 561 nm: dual band filter FITC/CY3
- Quad band filter FF01-440/521/607/700-25

8.2 DNA origami annealing and AFM measurements

Annealing. The bacterial M13mp18 genome (tilibit nanosystems) was folded with the help of thermal annealing into the triangular shape using staple strands as described in [73]. The following strands have been modified as follows (with the modification highlighted in red):

t7s18g	5'-Alexa647-TTCGGATGGCACGAGAATGACCATAATCGTTTACCAGACGAC
t-7s18g	5'-Alexa647-TTAAAACACTTAATCTTGACAAGAACTTAATCATTGTGAATT
t-1s28g	5'-Alexa647-TTTTTCACCAGCCTGGCCCTGAGAGAAAGCCGGCGAACGTGG
t-7s28g	5'-Alexa647-TTTTCCAGTCCTTATAAATCAAAAGAGAACCATCACCCAAAT
t4s7f	5'-Biotin-TTGGAGGGAATTTAGCGTCAGACTGTCCGCCTCC
t4s17f	5'-Biotin-TTGATTAGAGATTAGATACATTTCGCAAATCATA
t4s27f	5'-Biotin-TTTGACCTGACAAATGAAAAATCTAAAATATCTT

The annealing buffer consists of 1x TAE with 15 mM MgCl2 (Carl Roth). For purification, centrifugal filters (Amicon Ultra, 100K, Millipore) have been used. A detailed annealing protocol can be found in the appendix chapter 9.

AFM measurements. The DNA origami sample was adsorbed on freshly cleaved mica. The contained $MgCl_2$ in the sample buffer facilitates adsorption to the negatively-charged substrate. After incubation for around five minutes, the sample was rinsed with MQ water and blown dry with nitrogen. The images were recorded with a Bruker Multimode8 microscope in tapping-mode in air. Tap150AlG (Budget Sensors) cantilevers were used that feature a tip radius of below 10 nm, a 5 N/m force constant, and a 150 kHz resonance frequency.

8.3 Diatom sample preparations

Cell cultivation. *Thalassiosira pseudonana* (Hustedt) Hasle et Heimdal clone CCMP1335 was cultured in artificial seawater medium according to the North East Pacific Culture Collection protocol ("NEPC medium") at 18 °C under constant light at 5,000–10,000 lux as described previously [116].

tpSil3 expression vector. DNA sequences of the PCFPs with their codons optimized for expression in *T. pseudonana* [6] were synthesized by GeneArt with restriction sites added to allow for single step cloning into the diatom expression vectors (see table in appendix chapter 9). For cytosolic expression of the FPs, the plasmid DNA was digested with EcoRV and NotI and ligated into the EcoRV and NotI sites of pTpfcp30. For expression of the FPs as a C-terminal gene fusion with tpSil3 the plasmid DNA was digested with EcoRV and NotI and ligated into the EcoRV and NotI sites of pTpfcp/Sil3nt [21]. See Figure 8.1 for the vector scheme including the Sil3-fusion construct.

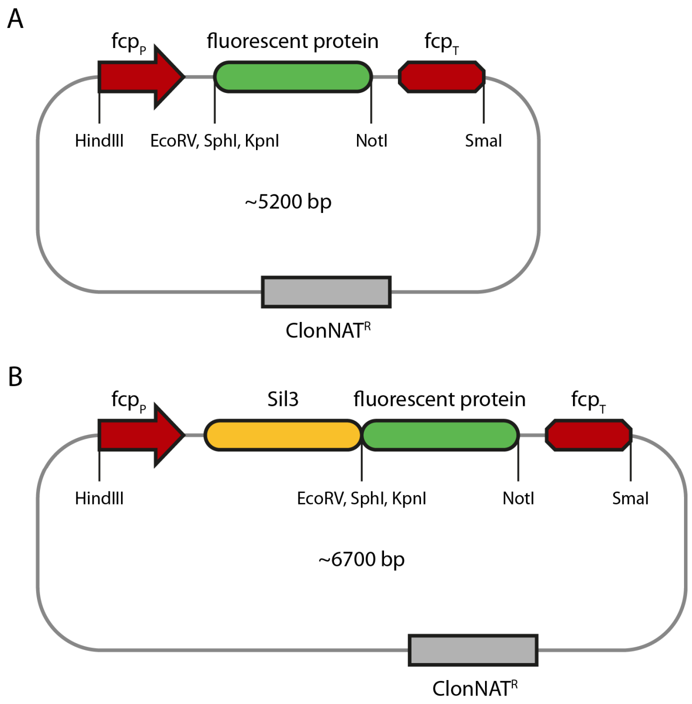


Figure 8.1 Vector design for cytosolic and Sil3 expression A) Cytosolic expression featuring the fluorescent protein in between the fcp promoter and terminator. B) Sil3 fusion constructs, with an N-terminal Silaffin3 followed by the C-terminal FPs.

Cingulin expression vector. Cingulins including native promoter and terminator have been cloned according to [18] by replacing the fcp promoter, assuming promoter and terminator lie within 1000 bp up- and downstream, respectively. For the two-color cingulin measurements in chapter 6, two vectors have been created similar to the tpSil3 expression vectors: CinY2 is C-terminally fused to eGFP and CinW2 is C-terminally fused to Dendra2.

Genetic transformation. The vector constructs were introduced into *T. pseudonana* cells using the Biorad PDS-1000/He particle delivery system as described previously [5]. Co-transformations were performed with the pTpfcp/nat plasmid DNA30 for selection of transformed cell lines on agar plates containing 150 µg/mL ClonNat (Nourseothricin, Jena Biosciences). Similarly, for the two-color cingulin construct, Zeocin™ (Phleomycin D1, InvivoGen) at 100 ng/ml has been chosen as a second antibiotic for selection in the second transformation step i.e. after CinY2-eGFP with ClonNat resistance has been introduced.

Biosilica isolation. Due to the intense chloroplast autofluorescence, it was necessary for PALM imaging of FPs fused to tpSil3 to remove cytosolic components via SDS and EDTA treatment as previously described [17]. A detailed protocol can be found in the appendix chapter 9. Cell walls were resuspended in PBS prior to imaging.

Microring isolation and immunostaining. After biosilica isolation, the biosilica has been dissolved using acidified ammonium fluoride solution as described in [17], [117]. The pelleted microrings are transferred onto poly-L-lysine coated coverslips and immunostained using GFPTrap®-Alexa647 (ChromoTek) as the fluorescent marker and Roti®-Block (Carl Roth) as a blocking agent. A detailed protocol can be found in the appendix chapter 9.

SEM imaging. For the SEM images shown throughout this work, diatoms were lysed (as described under Biosilica Isolation above), dehydrated with ethanol and afterwards critically point dried in a Leica-CPD 300. The dry diatoms were spread onto a carbon conductive adhesive tape and sputter coated with platinum in a Baltec SCD 050 (at 40 mA for 40 s). Imaging was performed using a JEOL JSM-7500F scanning electron microscope with an SE2 detector and an acceleration voltage of 15kV.

8.4 Fluorescence imaging conditions

Sample chamber. For all experiments except the DNA origami measurements presented in chapter 3, a simple 12-well slide and an ethanol-cleaned coverslip or in the case of the cingulin experiments a poly-L-lysine coated coverslip has been used. For the DNA origami measurements, a polyethylene glycol (PEG)-passivated quartz surface is coated with neutravidin to allow specific biotin binding according to [118].

***In vivo* diatom imaging.** For screening of Diatoms *in vivo* (chapters 4 and 6), the FPs have been imaged using the appropriate wavelengths for 100 ms at a power of 40 mW (end of fiber). In order to visualize the chloroplasts, a low 647 nm activation was used (1.5 mW for 100 ms). Photo-conversion for screening was performed by irradiating with UV light (405 nm) for 1 sec at 20 mW.

Diatom super-resolution imaging. For the imaging of Dendra2 & mEOS3.2 presented in chapter 5 and 6, a 488 nm excitation at 20 mW has been used to image pre-converted FPs for at least 1000 frames at 10 fps. For Dronpa, additional continuous UV (405 nm) illumination at 8 mW has been performed.

DNA origami super-resolution imaging. For the super-resolution imaging of DNA origami structures presented in chapter 3, Alexa647 has been excited using 647 nm at 100 mW at a frame rate of 10 fps. The UV (405 nm) photo-activation has been performed using an increasing activation power from 0 to 5 mW to adjust the localization count to an appropriate level by eye.

8.5 Buffer systems

***In vivo* diatom imaging.** For screening of Diatoms *in vivo* (chapter 4) and spinning disk microscopy, the standard NEPC medium was used.

PALM imaging. For PALM imaging of tpSil3 in isolated biosilica (chapter 5), PBS at pH 7.4 was used.

PALM-STORM on cingulins. For the studies in chapter 6, an oxygen scavenging system as well as a blinking buffer based on cysteamine had to be introduced into the buffer. The final buffer consists of:

- 100mM MEA (Cysteamine hydrochloride, M6500, Sigma Aldrich)
- 1% w/v Glucose
- 5 mg/ml Pyranose oxidase (Sigma Aldrich)
- 1 kU/ml Catalase (Calbiochem)

in PBS, pH 7.4.

Pyranose oxidase and Catalase are introduced right before measuring as the enzymes degrade over time.

STORM on DNA origami. For the DNA origami measurements (chapter 3), salts to stabilize the DNA origami have to be added to the buffer which is otherwise similar to the one described for PALM-STORM on cingulins above:

- 50mM Tris pH8
- 10mM NaCl
- 20mM $MgCl_2$
- 100mM MEA
- 1% w/v Glucose
- 5 mg/ml Pyranose oxidase
- 1 kU/ml Catalase

in MQ water.

9 APPENDICES

9.1 Tables and Protocols

CONTENTS

Table of codon usage in T. pseudonana 106

Table of codon optimized DNA sequences and amino acid sequence of all PCFPs used in this work 107

Protocol for biosilica isolation via cell lysis 108

Protocol for microring preparation via ammonium fluoride treatment of biosilica and immunostaining 109

Protocol for DNA origami folding 111

Table of codon usage in *T. pseudonana*

AmAcid	**Codon**	**Number**	**/1000**	**Frac**	**AmAcid**	**Codon**	**Number**	**/1000**	**Frac**
Gly	GGG	70178	11.53	0.17	Trp	TGG	77636	12.76	1
Gly	GGA	162724	26.74	0.4	End	TGA	24435	4.02	0.53
Gly	GGT	104470	17.17	0.25	Cys	TGT	60179	9.89	0.52
Gly	GGC	74032	12.17	0.18	Cys	TGC	55820	9.17	0.48
Glu	GAG	198991	32.7	0.53	End	TAG	10826	1.78	0.24
Glu	GAA	179974	29.58	0.47	End	TAA	10498	1.73	0.23
Asp	GAT	217000	35.66	0.61	Tyr	TAT	71391	11.73	0.44
Asp	GAC	139428	22.91	0.39	Tyr	TAC	90170	14.82	0.56
Val	GTG	118429	19.46	0.31	Leu	TTG	166242	27.32	0.32
Val	GTA	73099	12.01	0.19	Leu	TTA	42921	7.05	0.08
Val	GTT	109002	17.91	0.28	Phe	TTT	109901	18.06	0.52
Val	GTC	85404	14.04	0.22	Phe	TTC	99631	16.37	0.48
Ala	GCG	63766	10.48	0.14	Ser	TCG	87564	14.39	0.15
Ala	GCA	153325	25.2	0.35	Ser	TCA	113167	18.6	0.2
Ala	GCT	129924	21.35	0.3	Ser	TCT	108826	17.88	0.19
Ala	GCC	93160	15.31	0.21	Ser	TCC	88542	14.55	0.15
Arg	AGG	60977	10.02	0.17	Arg	CGG	30645	5.04	0.09
Arg	AGA	79990	13.15	0.23	Arg	CGA	65902	10.83	0.19
Ser	AGT	102178	16.79	0.18	Arg	CGT	79268	13.03	0.22
Ser	AGC	72866	11.97	0.13	Arg	CGC	37854	6.22	0.11
Lys	AAG	193898	31.87	0.62	Gln	CAG	91733	15.08	0.39
Lys	AAA	119576	19.65	0.38	Gln	CAA	143739	23.62	0.61
Asn	AAT	128412	21.1	0.48	His	CAT	81307	13.36	0.54
Asn	AAC	138500	22.76	0.52	His	CAC	68901	11.32	0.46
Met	ATG	148949	24.48	1	Leu	CTG	67141	11.03	0.13
Ile	ATA	51690	8.49	0.18	Leu	CTA	53160	8.74	0.1
Ile	ATT	124055	20.39	0.43	Leu	CTT	92179	15.15	0.18
Ile	ATC	111439	18.31	0.39	Leu	CTC	100123	16.45	0.19
Thr	ACG	82712	13.59	0.23	Pro	CCG	49675	8.16	0.17
Thr	ACA	107227	17.62	0.29	Pro	CCA	100190	16.47	0.35
Thr	ACT	100543	16.52	0.27	Pro	CCT	82575	13.57	0.29
Thr	ACC	75250	12.37	0.21	Pro	CCC	51624	8.48	0.18

Table of codon optimized DNA sequences and amino acid sequence of all PCFPs used in this work

PATagRFP

702 nucleotides, 234 amino acids

```
  1 ATGTCCGAGTTGATCAAAGAAAACATGCAC ATGAAGTTGTACATGGAAGGTACTGTGAAC AACCACCACTTCAAGTGCACATCAGAGGGC GAGGGAAAGCCATACGAGGGAACACAAACA ATGCGTATCAAGGTGGTGGAAGGTGGACCA
  1  M  S  E  L  I  K  E  N  M  H   M  K  L  Y  M  E  G  T  V  N   N  H  H  F  K  C  T  S  E  G   E  G  K  P  Y  E  G  T  Q  T   M  R  I  K  V  V  E  G  G  P
151 TTGCCATTCGCATTCGACATCTTGGCCACG AGCTTTATGTACGGATCCTCCACATTCATC AACCACACGCAGGGAATCCCAGACTTCTGG AAGCAATCATTCCCAGAGGGATTCACATGG GAGCGTGTGACAACATACGAGGATGGTGGT
 51  L  P  F  A  F  D  I  L  A  T   S  F  M  Y  G  S  S  T  F  I   N  H  T  Q  G  I  P  D  F  W   K  Q  S  F  P  E  G  F  T  W   E  R  V  T  T  Y  E  D  G  G
301 GTTTTGACGGCAACGCAAGACACCTCATTG CAAGACGGATGCTTGATCTACAACGTGAAG ATCAGGGGAGTGAACTTCCCATCAAACGGA CCAGTGATGAAGAAAAAGACGTTGGGATGG GAGCCATCCACCGAGAAGTTGAAGCCAGCA
101  V  L  T  A  T  Q  D  T  S  L   Q  D  G  C  L  I  Y  N  V  K   I  R  G  V  N  F  P  S  N  G   P  V  M  K  K  K  T  L  G  W   E  P  S  T  E  K  L  K  P  A
451 GATGGTGGATTGGAGGGAAGAGTGGATATG GCATTGAAGTTGGTTGGAGGTGGACACTTG ATCTGCAACTTCAAGACGACGTACCGTTCC AAGAAGCCAGCCAAGAACTTGAAGATGCCT GGTGTTTACTACGTGGACCGTCGTTTGGAG
151  D  G  G  L  E  G  R  V  D  M   A  L  K  L  V  G  G  G  H  L   I  C  N  F  K  T  T  Y  R  S   K  K  P  A  K  N  L  K  M  P   G  V  Y  Y  V  D  R  R  L  E
601 ATCATCAAAGAGGCCGACAAAGAGACATAC TGGGAGCAACATGAGGTGGCAGTGGCAAGA TACTCCGACTTGCCATCAAAGTTGGGACAC AAGTTGAACTGA
201  I  I  K  E  A  D  K  E  T  Y   W  E  Q  H  E  V  A  V  A  R   Y  S  D  L  P  S  K  L  G  H   K  L  N  *
```

PAmCherry

711 nucleotides, 237 amino acids

```
  1 ATGGTGTCAAAGGGTGAAGAGGACAACATG GCCATCATCAAAGAATTCATGCGTTTCAAG GTGCACATGGAAGGATCCGTGAACGGACAT GTGTTTGAGATTGAGGGTGAAGGCGAAGGA CGACCATACGAGGGAACACAGACAGCAAAG
  1  M  V  S  K  G  E  E  D  N  M   A  I  I  K  E  F  M  R  F  K   V  H  M  E  G  S  V  N  G  H   V  F  E  I  E  G  E  G  E  G   R  P  Y  E  G  T  Q  T  A  K
151 TTGAAAGTGACGAAGGGTGGACCATTGCCA TTCACGTGGGATATTTTGTCGCCACAGTTT ATGTACGGATCCAACGCATACGTGAAGCAC CCAGCAGATATTCCAGACTACTTCAAGTTG TCCTTCCCAGAGGGATTCAAGTGGGAGCGT
 51  L  K  V  T  K  G  G  P  L  P   F  T  W  D  I  L  S  P  Q  F   M  Y  G  S  N  A  Y  V  K  H   P  A  D  I  P  D  Y  F  K  L   S  F  P  E  G  F  K  W  E  R
301 GTGATGAAGTTTGAGGATGGTGGTGTTGTG ACCGTGACGCAGGATTCATCATTGCAAGAC GGTGAATTCATCTACAAAGTGAAGTTGCGT GGAACGAACTTCCCATCCGATGGACCAGTG ATGCAGAAAAAGACAATGGGATGGGAGGCC
101  V  M  K  F  E  D  G  G  V  V   T  V  T  Q  D  S  S  L  Q  D   G  E  F  I  Y  K  V  K  L  R   G  T  N  F  P  S  D  G  P  V   M  Q  K  K  T  M  G  W  E  A
451 TTGTCAGAGAGAATGTACCCAGAGGATGGT GCCTTGAAGGGTGAAGTTAAGCCACGTGTG AAGTTGAAAGACGGTGGACACTACGATGCA GAAGTCAAGACGACGTACAAGGCAAAGAAG CCAGTTCAGTTGCCTGGTGCATACAACGTG
151  L  S  E  R  M  Y  P  E  D  G   A  L  K  G  E  V  K  P  R  V   K  L  K  D  G  G  H  Y  D  A   E  V  K  T  T  Y  K  A  K  K   P  V  Q  L  P  G  A  Y  N  V
601 AACCGAAAGTTGGACATCACGTCCCACAAC GAGGATTACACGATCGTTGAGCAATACGAG AGAGCAGAGGGACGTCATAGTACAGGTGGA ATGGACGAGTTGTACAAATGA
201  N  R  K  L  D  I  T  S  H  N   E  D  Y  T  I  V  E  Q  Y  E   R  A  E  G  R  H  S  T  G  G   M  D  E  L  Y  K  *
```

PAGFP

720 nucleotides, 240 amino acids

```
  1 ATGGTGAGCAAGGGCGAGGAGCTGTTCACC GGGGTGGTGCCCATCCTGGTCGAGCTGGAC GGCGACGTAAACGGCCACAAGTTCAGCGTG TCCGGCGAGGGCGAGGGCGATGCCACCTAC GGCAAGCTGACCCTGAAGTTCATCTGCACC
  1  M  V  S  K  G  E  E  L  F  T   G  V  V  P  I  L  V  E  L  D   G  D  V  N  G  H  K  F  S  V   S  G  E  G  E  G  D  A  T  Y   G  K  L  T  L  K  F  I  C  T
151 ACCGGCAAGCTGCCCGTGCCCTGGCCCACC CTCGTGACCACCTTCAGCTACGGCGTGCAG TGCTTCAGCCGCTACCCCGACCACATGAAG CAGCACGACTTCTTCAAGTCCGCCATGCCC GAAGGCTACGTCCAGGAGCGCACCATCTTC
 51  T  G  K  L  P  V  P  W  P  T   L  V  T  T  F  S  Y  G  V  Q   C  F  S  R  Y  P  D  H  M  K   Q  H  D  F  F  K  S  A  M  P   E  G  Y  V  Q  E  R  T  I  F
301 TTCAAGGACGACGGCAACTACAAGACCCGC GCCGAGGTGAAGTTCGAGGGCGACACCCTG GTGAACCGCATCGAGCTGAAGGGCATCGAC TTCAAGGAGGACGGCAACATCCTGGGGCAC AAGCTGGAGTACAACTACAACAGCCACAAC
101  F  K  D  D  G  N  Y  K  T  R   A  E  V  K  F  E  G  D  T  L   V  N  R  I  E  L  K  G  I  D   F  K  E  D  G  N  I  L  G  H   K  L  E  Y  N  Y  N  S  H  N
451 GTCTATATCATGGCCGACAAGCAGAAGAAC GGCATCAAGGCCAACTTCAAGATCCGCCAC AACATCGAGGACGGCAGCGTGCAGCTCGCC GACCACTACCAGCAGAACACCCCCATCGGC GACGGCCCCGTGCTGCTGCCCGACAACCAC
151  V  Y  I  M  A  D  K  Q  K  N   G  I  K  A  N  F  K  I  R  H   N  I  E  D  G  S  V  Q  L  A   D  H  Y  Q  Q  N  T  P  I  G   D  G  P  V  L  L  P  D  N  H
601 TACCTGAGCCACCAGTCCAAGCTGAGCAAA GACCCCAACGAGAAGCGCGATCACATGGTC CTGCTGGAGTTCGTGACCGCCGCCGGGATC ACTCTCGGCATGGACGAGCTGTACAAGTAA
201  Y  L  S  H  Q  S  K  L  S  K   D  P  N  E  K  R  D  H  M  V   L  L  E  F  V  T  A  A  G  I   T  L  G  M  D  E  L  Y  K  *
```

mEOS3.2

681 nucleotides, 227 amino acids

```
  1 ATGTCTGCAATCAAGCCAGACATGAAGATC AAGTTGCGTATGGAAGGCAACGTGAACGGA CACCATTTCGTGATCGACGGTGATGGAACT GGAAAGCCATTCGAGGGAAAGCAGTCCATG GATTTGGAAGTGAAAGAGGGTGGACCATTG
  1  M  S  A  I  K  P  D  M  K  I   K  L  R  M  E  G  N  V  N  G   H  H  F  V  I  D  G  D  G  T   G  K  P  F  E  G  K  Q  S  M   D  L  E  V  K  E  G  G  P  L
151 CCATTCGCATTCGACATCTTGACAACGGCC TTCCATTACGGAAACCGTGTGTTCGCAAAG TACCCAGACAACATCCAGGACTACTTCAAG CAGTCCTTCCCAAAGGGATACTCCTGGGAG AGATCATTGACATTCGAGGACGGTGGAATC
 51  P  F  A  F  D  I  L  T  T  A   F  H  Y  G  N  R  V  F  A  K   Y  P  D  N  I  Q  D  Y  F  K   Q  S  F  P  K  G  Y  S  W  E   R  S  L  T  F  E  D  G  G  I
301 TGCAACGCACGTAACGATATTACGATGGAA GGCGACACGTTCTACAACAAAGTGCGATTC TACGGAACGAACTTCCCAGCAAACGGACCA GTGATGCAGAAAAAGACATTGAAGTGGGAG CCATCCACCGAGAAGATGTACGTGCGTGAT
101  C  N  A  R  N  D  I  T  M  E   G  D  T  F  Y  N  K  V  R  F   Y  G  T  N  F  P  A  N  G  P   V  M  Q  K  K  T  L  K  W  E   P  S  T  E  K  M  Y  V  R  D
451 GGTGTCTTGACGGGTGATATTGAGATGGCC TTGTTGTTGGAGGGAAACGCACATTACCGT TGCGATTTCCGTACGACGTACAAGGCCAAA GAAAAGGGTGTTAAGTTGCCAGGTGCACAC TTCGTGGATCACTGCATCGAGATCTTGTCC
151  G  V  L  T  G  D  I  E  M  A   L  L  L  E  G  N  A  H  Y  R   C  D  F  R  T  T  Y  K  A  K   E  K  G  V  K  L  P  G  A  H   F  V  D  H  C  I  E  I  L  S
601 CACGACAAGGACTATAACAAAGTGAAGTTG TACGAGCACGCAGTGGCACACTCAGGATTG CCAGATAACGCAAGACGTTGA
201  H  D  K  D  Y  N  K  V  K  L   Y  E  H  A  V  A  H  S  G  L   P  D  N  A  R  R  *
```

Dendra2

693 nucleotides, 231 amino acids

```
  1 ATGAACACACCTGGAATCAACTTGATCAAA GAAGATATGCGAGTCAAGGTGCACATGGAA GGCAACGTGAACGGACATGCATTCGTGATT GAGGGCGAGGGAAAGGGAAAACCATACGAG GGAACACAGACGGCCAACTTGACAGTGAAA
  1  M  N  T  P  G  I  N  L  I  K   E  D  M  R  V  K  V  H  M  E   G  N  V  N  G  H  A  F  V  I   E  G  E  G  K  G  K  P  Y  E   G  T  Q  T  A  N  L  T  V  K
151 GAGGGTGCACCATTGCCATTCTCGTACGAC ATTTTGACAACGGCCGTGCATTACGGAAAC CGTGTGTTTACAAAGTACCCAGAGGACATC CCAGACTACTTCAAGCAATCATTCCCAGAG GGATACTCGTGGGAGCGTACAATGACATTC
 51  E  G  A  P  L  P  F  S  Y  D   I  L  T  T  A  V  H  Y  G  N   R  V  F  T  K  Y  P  E  D  I   P  D  Y  F  K  Q  S  F  P  E   G  Y  S  W  E  R  T  M  T  F
301 GAGGATAAGGGAATCTGCACGATCCGTTCC GACATTTCATTGGAGGGCGACTGTTTCTTC CAGAACGTGCGTTTCAAGGGAACGAACTTT CCACCAAACGGACCAGTGATGCAGAAAAAG ACATTGAAGTGGGAGCCATCCACCGAGAAG
101  E  D  K  G  I  C  T  I  R  S   D  I  S  L  E  G  D  C  F  F   Q  N  V  R  F  K  G  T  N  F   P  P  N  G  P  V  M  Q  K  K   T  L  K  W  E  P  S  T  E  K
451 TTGCATGTTCGTGACGGTTTGTTGGTCGGA AACATCAACATGGCCTTGTTGTTGGAGGGT GGTGGACACTACTTGTGCGATTTCAAGACG ACGTACAAGGCCAAGAAGGTGGTCCAATTG CCAGATGCACATTTCGTGGATCACCGTATC
151  L  H  V  R  D  G  L  L  V  G   N  I  N  M  A  L  L  L  E  G   G  G  H  Y  L  C  D  F  K  T   T  Y  K  A  K  K  V  V  Q  L   P  D  A  H  F  V  D  H  R  I
601 GAGATCTTGGGAAACGACTCCGACTACAAC AAAGTGAAGTTGTACGAGCACGCAGTGGCC AGATACTCTCCATTGCCATCACAAGTCTGG TGA
201  E  I  L  G  N  D  S  D  Y  N   K  V  K  L  Y  E  H  A  V  A   R  Y  S  P  L  P  S  Q  V  W   *
```

Dronpa

678 nucleotides, 226 amino acids

```
  1 ATGGTGTCAGTGATTAAGCCAGACATGAAG ATCAAGTTGCGTATGGAAGGTGCAGTGAAC GGACATCCATTCGCAATTGAGGGTGTTGGA TTGGGAAAGCCATTCGAGGGAAAGCAGTCG ATGGACTTGAAAGTGAAAGAGGGTGGACCA
  1  M  V  S  V  I  K  P  D  M  K   I  K  L  R  M  E  G  A  V  N   G  H  P  F  A  I  E  G  V  G   L  G  K  P  F  E  G  K  Q  S   M  D  L  K  V  K  E  G  G  P
151 TTGCCATTCGCCTACGATATTTTGACGACG GTGTTCTGCTACGGAAACCGTGTGTTCGCA AAGTACCCAGAGAACATCGTGGACTACTTC AAGCAGTCATTCCCAGAGGGATACTCGTGG GAGAGATCAATGAACTACGAGGACGGTGGA
 51  L  P  F  A  Y  D  I  L  T  T   V  F  C  Y  G  N  R  V  F  A   K  Y  P  E  N  I  V  D  Y  F   K  Q  S  F  P  E  G  Y  S  W   E  R  S  M  N  Y  E  D  G  G
301 ATCTGCAACGCAACAAACGATATTACGTTG GACGGTGACTGCTACATCTACGAGATCCGT TTCGATGGTGTCAACTTCCCAGCAAACGGA CCAGTGATGCAAAAGCGTACAGTGAAGTGG GAGCCATCCACAGAGAAGTTGTACGTGCGT
101  I  C  N  A  T  N  D  I  T  L   D  G  D  C  Y  I  Y  E  I  R   F  D  G  V  N  F  P  A  N  G   P  V  M  Q  K  R  T  V  K  W   E  P  S  T  E  K  L  Y  V  R
451 GATGGTGTCTTGAAGGGTGACGTGAACATG GCATTGTCATTGGAGGGTGGTGGACACTAC CGTTGCGATTTCAAGACGACGTACAAGGCA AAGAAGGTGGTCCAATTGCCAGACTACCAT TTCGTGGATCACCACATCGAGATCAAGTCC
151  D  G  V  L  K  G  D  V  N  M   A  L  S  L  E  G  G  G  H  Y   R  C  D  F  K  T  T  Y  K  A   K  K  V  V  Q  L  P  D  Y  H   F  V  D  H  H  I  E  I  K  S
601 CACGACAAGGACTACTCCAACGTGAACTTG CATGAGCACGCAGAGGCACATTCAGAGTTG CCACGTCAAGCCAAATGA
201  H  D  K  D  Y  S  N  V  N  L   H  E  H  A  E  A  H  S  E  L   P  R  Q  A  K  *
```

Protocol for biosilica isolation via cell lysis

General remarks:

- perform all washing steps in a 50 ml tube with 15-25 ml washing volume
- always spin the cells down at 3200 x g for 10 min

Starting material:

- *Thalassiosira pseudonana*, wet cells at a high density (10^6 cells/ml)
- or liquid culture (at least 100ml – spin down in 50ml tubes and pool pellets)

Procedure:

- lyse cells using 2% SDS, 100 mM EDTA at 60 °C for 60 min in a shaker, spin down
- wash the pellet with 100 mM EDTA by vortexing, spin down
- if necessary repeat the last 2 steps
- wash the pellet with MQ water by vortexing, spin down (1x)
- wash the pellet with acetone by vortexing, spin down (3x)
- wash the pellet with MQ water by vortexing, spin down (3x)
- now the material is ready for imaging
- can be stored in the fridge for several months when 10mM EDTA and 1mM PMSF is added for stabilization

Protocol for microring preparation via ammonium fluoride treatment of biosilica and immunostaining

Poly-L-Lysine slide preparation

- prepare the Sigma-Aldrich P8920 solution according to protocol: dilute solution 1:10 with MQ water, incubate with slides for 5 min, drain slides and dry at 60 °C for 1 h

Antibody labeling (if necessary):

- To label the GFPTrap® (Chromotek) with Alexa647, the following protocol has been carried out prior to immunostaining
- Reaction buffer:
    - 50 µl 80 µM Antibody (= 40 µM)
    - 2.5 µl 8 mM Alexa Fluor (= 200 µM)
    - 10 µl 1 M NaHCO3 (= 100 mM)
    - Add PBS to 100 µl
- Incubate reaction buffer for 1 h at room temperature in the dark
- Prepare Bio-Spin 6 (Bio Rad) purification column
- Exchange the buffer in the column to PBS
- Purify antibody using the Bio-Spin 6 column (take less elution buffer than indicated in the protocol to reduce the risk of unbound fluorophores)
- Take absorption spectra to determine dye/protein ratios. The ratio should be >0.5
- Store in the dark at 4 °C

Diatom starting material:

- Biosilica (*T. pseudonana* cells, cleaned cell walls, around 10^6 cells/ml))

NH_4F treatment:

- Dilute cells in an Eppendorf vial; use around 800 µl MQ water and 200 µl concentrated cells, giving a max. volume of 1 ml in total
- Spin down (5 min at 10000 x g), and remove the supernatant by careful pipetting
- Prepare NH_4F buffer:
    - 1 ml 10 M NH_4F
    - add 6 M HCl until pH is around 4.5 (requires around 300 µl of 6 M HCl)
- Add at least 750 µl of NH_4F buffer to the diatoms
- Incubate 1 h at room temperature
- Spin down, discard NH_4F
- Rinse with MQ water by carefully pipetting up and down
- add MQ water, vortex, spin down, remove MQ water

Immunostaining:

- Add blocking solution to a maximum of 1 ml of diatoms:
    - dilute Roti-Block solution 1:10
    - add 400 µl Roti-Block 0.05% Tween20
- Put a droplet (around 200 µl) on Poly-L-Lysine slide
- Incubate 1 h for blocking and immobilization, cover with a wet tissue to avoid drying out
- Wash extensively in PBS / MQ water by shaking the slide with tweezers under water
- Prepare the primary antibody solution:
    - Add antibody 1:100
    - add again 400 µl Roti-Block and 0.05% Tween20
- Incubate 1 h at room temperature
- Wash extensively (as noted before)

Protocol for DNA origami folding

Notes:

- This protocol results in around 30 µl of 20 - 40 nM origami solution
- Dilute all staple strands to 3 µM and the scaffold to 100 nM for better handling
- Assuming a master-mix (of all strands that are often used and constant) of 200 staples: mixing those staples at concentrations of 100 µM will results in a final concentration of 500 nM per strand
- In the final solution, master-mix staples will account for 30 µl in 100 µl – so the final concentration for all strands is reduced to 150 nM per strand

Mixing final concentration:

- 30 µl master-mix = 150 nM per strand
- 5 µl per modification = 150 nM, dilute stock to 3 µM
- 5 µl M13mp18 scaffold = 5 nM, dilute stock to 100 nM
- 10 µl 10 x TAE = 1 x TAE
- 1.5 µl MgCl2 = 15 mM, 1 M Stock
- fill with MQ water to 100 µl

Annealing (PCR):

- starting at 80 °C cool down in 0.5 °C steps (each 12 s) to 55 °C (= 50 steps)
- cool down in 0.3 °C steps (each 48 s) to room temperature (99 steps)
- hold at 4 °C
- the whole procedure takes around 1.5 h

Spin-filtration (Amicon Ultra-0.5 mL, 100 kDa MWCO):

- mix 200 µl 1 x TAE 10 mM MgCl2 and 100 µl sample
- spin 10 min at 10000 x g
- remove supernatant
- add 300 µl 1 x TAE 10 mM MgCl2
- spin 10 min at 10000 x g
- turn around and spin 5 min at 5000 x g to empty the filter

9.2 Satellite projects

CONTENTS

Quantitative fluorescence intensity analysis of 3D time-lapse confocal microscopy data in diatoms 112

Applying neural networks to filter SMLM localizations 118

In vivo imaging at super-resolution conditions 121

Quantifying chromatic aberrations in the microscope using fiducials 123

9.2.1 Quantitative fluorescence intensity analysis of 3D time-lapse confocal microscopy data in diatoms

For a collaboration with the research group of Prof. Nils Kröger (B CUBE, TU Dresden) I developed image analysis scripts to automate fluorescence quantification for 3D confocal microscopy data which will be presented in this chapter. This project was inspired by the discovery of the membrane protein Silicanin1 (Sin1) in *T. pseudonana* with unknown function as a component of SDVs. The corresponding publication is currently under review (present author list: Alexander Kotzsch, Philip Gröger, Damian Pawolski, Paul H. H. Bomans, Nico A. J. M. Sommerdijk , Michael Schlierf, and Nils Kröger) and will be referred to by reference [119]. Note that this chapter will - unlike the aforementioned publication - not focus on Sin1 and its function but on the analysis routine that has been established by myself.

In order to elucidate the function of Sin1, time-lapse confocal microscopy has been applied to visualize GFP-tagged Sin1 *in vivo* during cell division. Therewith, the fluorescence signal of GFP can be monitored in three dimensions by acquiring multiple Z-stacks of individual diatom cells over time. A quantitative analysis of this signal was necessary to associate changes of fluorescence over time to certain compartments of the diatom (i.e. cytosol, plasma membrane, and mid-cell region where the cell cleavage occurs). With this time-dependent association, detailed conclusions concerning the function of Sin1 during different stages of the cell cycle are possible. Simultaneously, PDMPO, a fluorescent silica tracer is recorded in a second channel to monitor silica biogenesis. As the analysis for this channel is the same as for to the GFP channel, the PDMPO channel will not be shown.

The raw data for a single diatom cell consist of a Z-stack of nine slices spaced 0.5 µm apart, taken every 3.5 minutes, with a total of 100 image stacks (corresponding to almost 6 hours). The cell is imaged with its cylinder axis parallel to the surface (Figure 9.1A). In ImageJ the cell is rotated to be aligned vertically, cropped, and the background fluorescence signal set to zero (by subtracting the minimum intensity value from the whole image stack). For the fluorescence analysis, a sum intensity projection of the Z-stack will be used. As the cylindrical 3D shape would introduce errors in this projection (i.e. fluorescence in the plasma membrane region would overlap with cytoplasmic signal), only the fluorescence from the central three slices (where the diatom shape is mostly perpendicular to the z-slice) is summed up. This modification will be compensated for later on in the analysis process. The resulting fluorescence patterns for four points in time during cell division can be seen in Figure 9.1B and C.

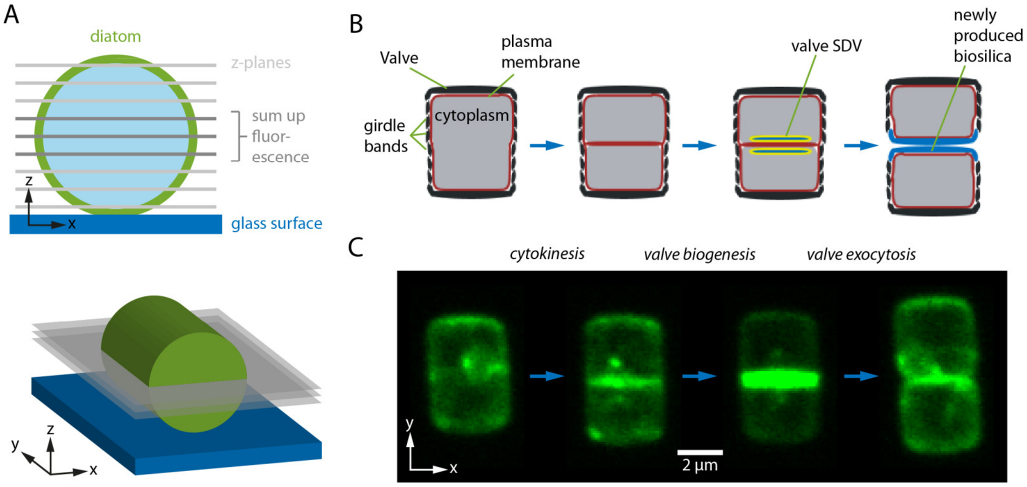


Figure 9.1 Image acquisition illustrations and time-lapse fluorescence images A) Schematic view along the cylinder axis (top) and as a 3D representation (bottom) of a centric diatom at the glass surface during imaging. The nine confocal Z-planes are shown and the central three planes are highlighted. B) Schematic representation of the part of the diatom cell cycle that has been recorded. C) Corresponding Sin1-GFP fluorescence images. Panels B and C have been adapted from [119].

Further processing of the image stack is carried out in MATLAB. In order to correct for sample or stage drift during the long acquisition time, a cross-correlation based image registration is applied [120]. Figure 9.2 shows a kymograph (i.e. a projection in the x-direction for each frame) before and after registration. The motion of the sample is displayed by noise in the y-direction of the kymograph (Figure 9.2A). This motion is greatly reduced after the image correction, resulting in a smooth kymograph (Figure 9.2B). Additionally, the diatom growth over time can be seen in this kymograph as an increase of the fluorescence distribution in the y-direction over time.

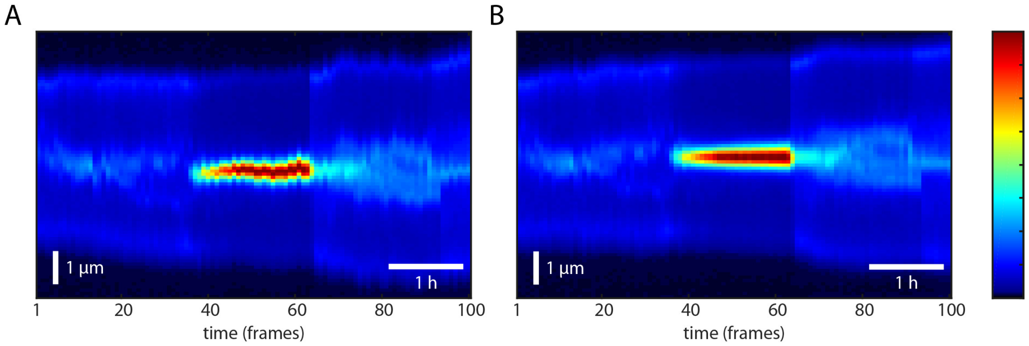


Figure 9.2 Image registration to correct sample drift A) Kymograph (projection of each frame in x-direction) of Sin1-GFP before drift correction during cell division. B) Kymograph after cross-correlation based image registration displays greatly reduced motion.

For the fluorescence quantification itself, the stabilized images (as shown in Figure 9.1C) have to be divided into regions containing the different cell compartments. In Figure 9.3A and B, the rectangular sub-regions that split the cell into plasma membrane, cytoplasm, and mid cell region are displayed and annotated. As the cell grows during the time of recording and slight surface movements might still be present, it is important to adjust the sub-region sizes accordingly. By manually defining the sub-regions for four key frames and interpolating the size in between, the cell growth can be accounted for (Figure 9.3D). The fluorescence intensity for each frame and each sub-region is then summed up and plotted against time (Figure 9.3C). The resulting images show clearly distinct patterns for each cell compartment.

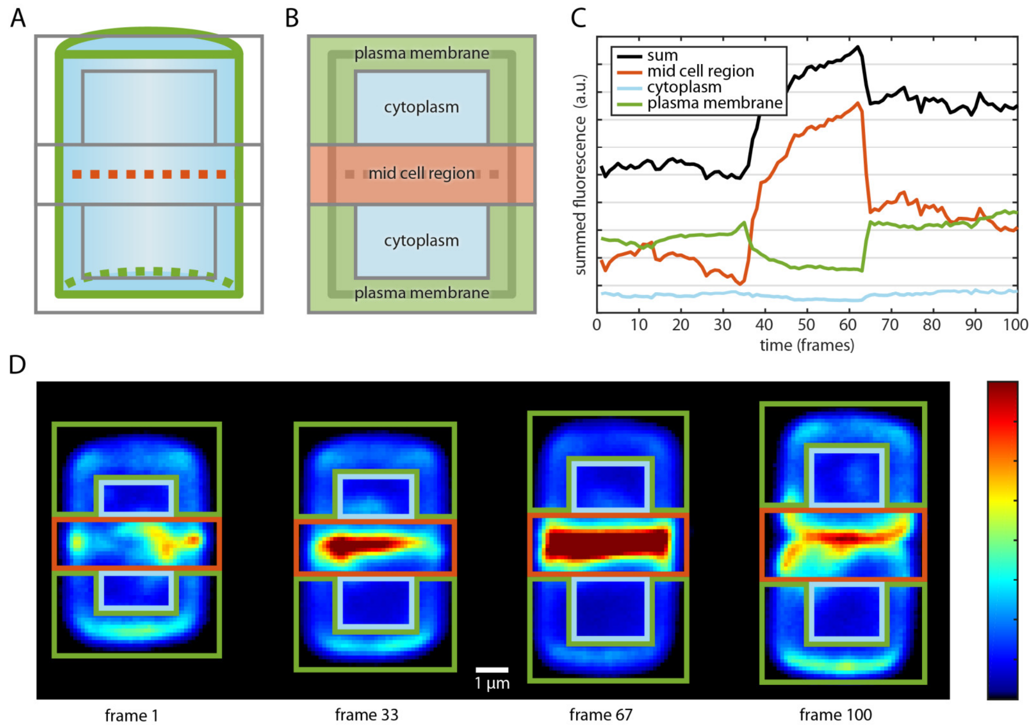


Figure 9.3 Fluorescence quantification using rectangular areas A) Schematic view of the diatom in imaging direction. Plasma membrane (green), cytoplasm (blue), and mid cell region (red) are highlighted as well as the rectangular areas for fluorescence summation (gray) are shown. B) Same view as in panel B with the rectangular analysis areas highlighted. C) The summed up fluorescence per rectangle is plotted over time. D) To account for cell growth, the rectangular areas are defined for four points in time of the image stack and interpolated for points in time in between.

The shown graph has already been corrected to account for fluorescence above or below the three central Z-planes. The following paragraph will detail the mathematical transformations that extrapolate the fluorescence signal to represent a perfect cylindrical diatom cell. This includes the reasonable assumption that the fluorescence is evenly distributed across the respective regions.

The rectangular regions are defined by the diatom length L, the diatom radius R, the thickness of the plasma membrane D_{PM}, as well as the thickness of the mid cell region D_{MID} (Figure 9.4A). For a perfectly cylindrical “model diatom” the following volumes can be calculated for the three regions (cytoplasm CYT', mid cell region MID', and plasma membrane PM') that all represent a cylinder by themselves (Figure 9.4B):

$$CYT' = \pi \cdot \underbrace{(R - D_{PM})^2}_{\text{cytoplasm cylinder}} \cdot \underbrace{(L - D_{MID} - 2 \cdot D_{PM})}_{\text{reduced length}}$$

$$MID' = \underbrace{A}_{\pi \cdot R^2} \cdot D_{MID}$$

$$PM' = A \cdot L - CYT' - MID'$$

In comparison, the measured volumes are additionally restricted by the height of the three central Z-planes Z. First, the areas in the axial direction (the circular cylinder base) have to be defined (Figure 9.4C). The plasma membrane area A_{CYT} is a simple cuboid:

$$A_{CYT} = (R - D_{PM}) \cdot 2 \cdot Z$$

For the mid cell area, each of the two circular segments A_{CS} that are not occupied by the z-planes (i.e. being above or below the z-planes) are defined as:

$$A_{CS} = R^2 \cdot \cos^{-1}\left(\frac{Z}{2 \cdot R}\right) - \frac{Z}{2} \cdot \sqrt{R^2 - \left(\frac{Z}{2}\right)^2}$$

Now, the mid cell area A_{MID} can be calculated by subtracting A_{CS} from the complete circle area A:

$$A_{MID} = A - 2 \cdot A_{CS}$$

The plasma membrane in the central area $A_{PM,C}$ is simply the difference between the other two areas:

$$A_{PM,C} = A_{MID} - A_{CYT}$$

To create the final volumes, simple extrusion of these area according to Figure 9.4A can be performed:

$$PM = \underbrace{A_{PM,C} \cdot (L - D_{MID} - 2 \cdot D_{PM})}_{\text{central volume}} + \underbrace{A \cdot D_{PM}}_{\text{cap volume}}$$

$$CYT = A_{CYT} \cdot (L - D_{MID} - 2 \cdot D_{PM})$$

$$MID = A_{MID} \cdot D_{MID}$$

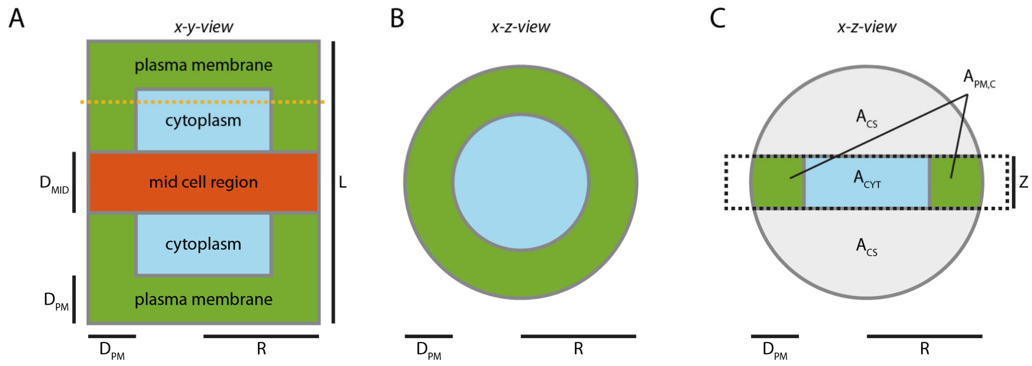


Figure 9.4 Geometric definitions for the correction factor calculation A) Schematic view of the diatom imaging areas in imaging direction. Slice position for panel B and C is indicated by a yellow dotted line. B) Axial view for a "perfect" model diatom. C) Same as B for the actual fluorescence analysis. Here, only fluorescence from the central region is picked up.

By dividing the volumes of plasma membrane, mid-cell and cytoplasma regions occupied in the entire "perfect" cell with the volumes occupied by the three central z-planes provides the desired correction factors F:

$$F_{PM} = PM'/PM$$

$$F_{MID} = MID'/MID$$

$$F_{CYT} = CYT'/CYT$$

With the developed analysis routine, the fluorescence signal of a cylindrical diatom shape can be effectively assigned to different cell compartments. By analyzing only Z-slices close to the center of the diatom cell, artifacts in the image projections could be circumvented. Correction factors could be defined that account for the missing fluorescence. The established analysis can easily extended to quantify fluorescence in cell compartments of different geometry.

9.2.2 Applying neural networks to filter SMLM localizations

When working with SMLM datasets, data filtering is often necessary to remove faulty localization created by an erroneous spot finding procedure or unspecific labeling. E.g. for datasets with a low signal to noise ratio, it is necessary to increase the sensitivity of the spot finding algorithm which may therefore introduce noise. Multiple strategies can be combined to perform this filtering:

- Threshold filtering based on the localization parameters given by the spot finding program. In the case of ThunderSTORM, parameters like fluorophore intensity, width of the fitted Gaussian, background intensity, or goodness of the Gaussian fit are accessible. With the developed SiMoNa GUI, exploring these parameters and using them as filter criteria is significantly simplified (chapter 2.7).
- Various additional post-processing steps can be performed that yield further filter criteria. Voronoi tessellation is one example and has been shown to be able to remove noise effectively in chapter 2.3.
- Finally, the manual selection of the structure in a region of interest in the image is possible.

All these procedures are both time consuming and have to be evaluated for each experiment individually. Therefore it is desirable to establish a completely automated, non-biased filtering method.

Neural networks are powerful tools for classification tasks, if a sufficiently large training dataset is available. With that in hand, it is possible to have the network to learn connections between input and output parameters. After a learning phase, the network is capable of classifying unknown samples (i.e. localizations) unsupervised and instantaneous. In the specific case of SMLM datasets, the input consists of the available parameters for each localization and the output would be a binary classification of whether a localization is filtered out or not.

This approach has been applied to cingulin datasets (i.e. CinW2-Dendra2, see chapter 6). Localizations from multiple datasets have been pooled and act as a training dataset for the network. The following eight localization parameters are taken into account as input:

- Width and offset of the fitted Gaussian
- Goodness of the Gaussian fit (Chi-square value)
- Standard deviation of the background around the localization
- Fluorophore intensity

- Localization precision
- Distance to the nearest neighboring localization (as a degree of local density)
- Area of the voronoi polygon around the localization (as described in chapter 2.3)

It should be noted that the x and y position of the localizations are not taken into account, as this information is strictly sample-dependent and does not reflect whether or not a localization has to be filtered out or not (unless e.g. localizations at the edge of the camera field of view shall always be filtered out). The output is defined as a "filter value" that ranges from zero (localization is filtered out) to one (localization is not filtered out). For the training dataset, this filter value has been determined as binary value by applying the Voronoi tessellation filtering as described and by manually selecting regions of microrings in the SMLM images to completely remove unspecific localizations.

The chosen network architecture is a feedforward network (as supplied by the MATLAB function `fitnet`) with an input and output layer of size 8 and 1, respectively, as dictated by the dimensionality of the input and output parameter. One hidden layer containing 20 nodes was determined to be sufficient, as an increase in network size did not improve results. See Figure 9.5D for a graphical representation of the network.

Figure 9.5A shows a detail of a CinW2-Dendra2 SMLM microring image. It is colored by the aforementioned binary filter value that has been assigned to each localization using manual filtering. After training on this and other datasets, the network can now assign the filter value to each localization autonomously. In Figure 9.5B, the resulting filter values are displayed for the example region. The continuous distribution can be interpreted as the probability that a certain localization will be kept and not filtered out. By setting a discrete threshold value, e.g. 0.7 (Figure 9.5C), it can be seen that the classification is extremely close to the original distribution. A histogram of the classification error (as defined by the difference between the target binary filter values and the filter values obtained from the network) displays a sharp peak around zero showing very little deviation from the original filter values (Figure 9.5E). A final test for the networks' performance can be obtained by running it on an unknown sample (that has not been part of the training data). Figure 9.5F shows again a part of a CinW2-Dendra2 microring pattern where the network effectively filtered out unspecific localizations.

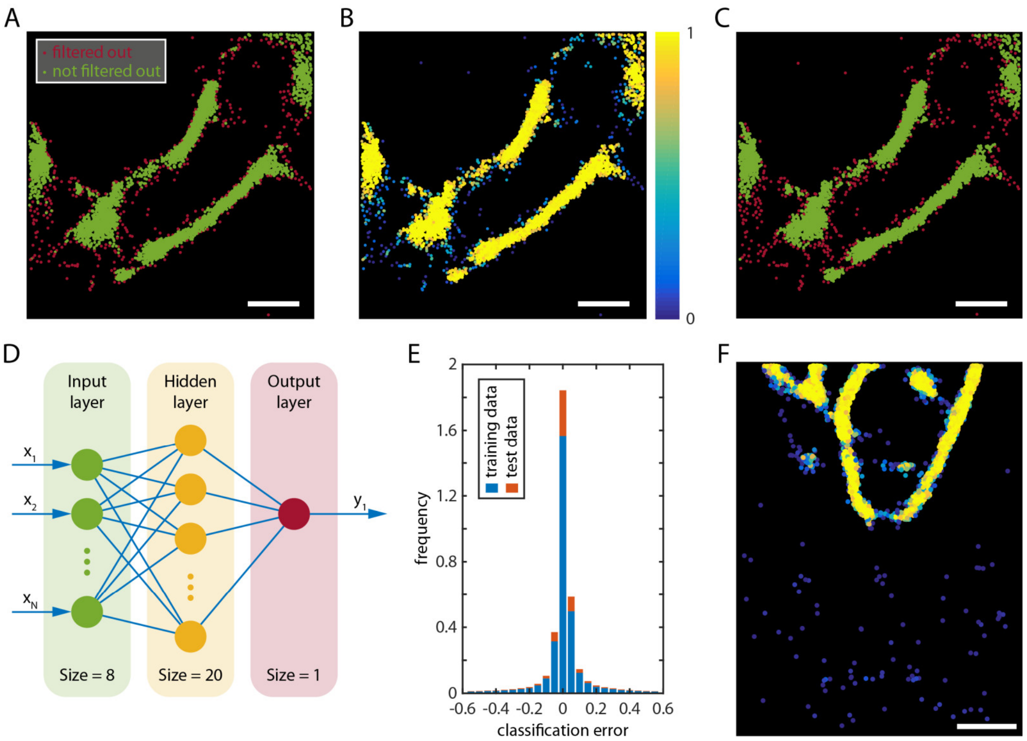


Figure 9.5 Network performance for CinW2-Dendra2 cingulin data filtering A) Scatter plot of a CinW2-Dendra2 filament, colored by the filter value obtained by manual filtering. B) Same region of interest as panel A), but filtered according to the networks' filter value. C) Same as panel B), but with a filter threshold set to 0.7 which results in a similar filter pattern as panel A). D) Feedforward Network architecture used for this study featuring one hidden layer containing 20 nodes. E) Histogram for the classification error (target value subtracted by network output) for the used training and test data. F) For an unknown sample, the network displays effective filtering as in panel B). All scale bars are 1 µm

These first exploratory experiments show the effectiveness of neural networks for classification tasks that will allow unsupervised high-throughput filtering of SMLM datasets. This idea can be easily extended to train networks to specifically detect multiple features of the dataset (i.e. to warn the user if a dataset displays strong drift, contains unspecific localizations, or to inform which areas of the field of view actually contain the sample). Thus, this approach would simplify and merge many post-processing steps.

9.2.3 *In vivo* imaging at super-resolution conditions using SOFI

Obtaining *in vivo* SMLM images was mainly hampered due to the chloroplasts autofluorescence that created a high background signal. Thus, it is highly desirable to establish a technique to reduce the chloroplast signal – not only for super-resolution imaging, but for fluorescence imaging in diatoms in general.

Super-resolution optical fluctuation imaging (SOFI) [114] might achieve these tasks. This method relies on statistical analysis of temporal fluctuations (e.g. blinking of fluorescence signal) and follows this basic workflow:

1. The structure of interest has to be fluorescently labeled with a blinking or more general a fluctuating probe. This fluctuation does not need to be photo-controllable.
2. A movie of the sample has to be recorded (as for SMLM).
3. For each pixel value, the *n*th-order cumulant of the original time series is taken.

The resulting SOFI image of *n*th-order features a decreased background and a theoretical resolution increase of $\sqrt{n}$. A detailed explanation for this behavior can be found in the original publication and will be covered here only briefly: An emitter positioned exactly at a certain pixel position will give rise to a highly fluctuating time series. Therefore, the auto-correlation curve for this pixel will display high correlation values. The *n*th-order cumulant is related to the *n*th-order correlation function and will – generally spoken – filter a signal based on the fluctuations in such a manner that only highly correlated fluctuations remain. This explains the background-reducing property of the SOFI algorithm, as a constant background (with fluctuations on a different timescale than those of the fluorescent probes) will be filtered out by the cumulant. The resolution gain can be explained as follows: Emitters at a certain pixel position will also introduce fluctuations in neighboring pixels (as the point-spread function is normally bigger than a single pixel), although to a lower extend. As the signal contribution to the neighboring pixel will yield nonlinearly lower correlation values, the cumulant will filter out the intensity contributions from neighboring emitters as well. Thereby, only the biggest contributions from the closest emitters remain, and thus the resolution is increased.

With these features in mind, it seems possible that SOFI might reduce chloroplast background in standard confocal diatom images *in vivo* as well. Figure 9.6A shows tpSil3 tagged with Dendra2 in the green channel (488 nm excitation) and chloroplast emission mainly in the red channel (647 nm excitation). SOFI was applied to 100 frames recorded in the green channel. A maximum intensity projection of these 100 frames in the left panel of Figure 9.6B clearly shows a strong chloroplast background which prevented single-molecule localization and hampered spotting the Dendra2 fluorescence in the cell wall. After applying the 4th-order SOFI algorithm, the chloroplast signal is greatly reduced, leaving the Dendra2 fluorescence behind (Figure 9.6B, right panel). A line scan quantifies the reduction of chloroplast autofluorescence as well as the background in general and visualizes peaks of tpSil3-Dendra2 fluorescence in the frustule region (Figure 9.6C). The resolution-gain is minimal, potentially due to sub-optimal blinking behavior of Dendra2. Potentially, a smaller pixel size and the application of advanced SOFI algorithms [121], [122] might increase the usability of super-resolution imaging *in vivo*.

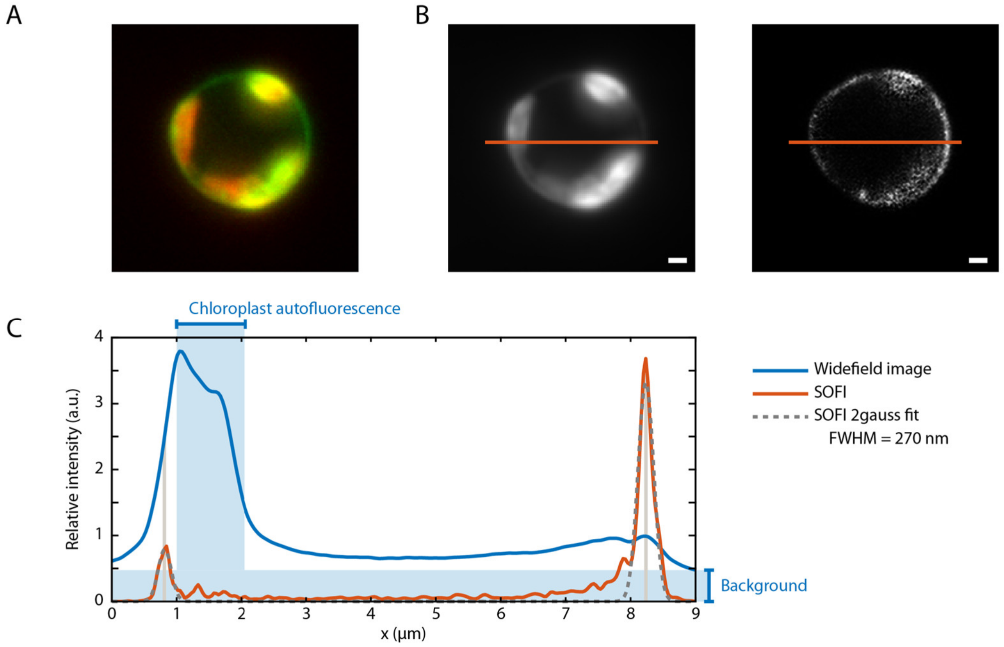


Figure 9.6 SOFI on tpSil3-Dendra2 in live cells A) Confocal fluorescence image of tpSil3-Dendra2 in the green channel (488 nm excitation) as well as chloroplast emission seen mainly in the red channel (647 nm excitation). B) Left panel: Maximum intensity projection of 100 frames recorded under 488 nm excitation showing strong chloroplast background. Right panel: 4th order SOFI algorithm applied to the image stack. Scale bars are 1 µm. C) Line scan of the area specified in panel B). Regions of chloroplast autofluorescence and background have been highlighted. A two Gaussian fit displays the position of tpSil3-Dendra2 fluorescence.

9.2.4 Quantifying chromatic aberrations in the microscope using fiducials

For the multicolor super-resolution imaging that has been performed in the cingulin studies presented in chapter 6, chromatic aberrations and thus offsets between the individual colors are of concern, especially if the degree of co-localization is of major interest. Therefore, chromatic aberrations have been quantified using fluorescent beads. The analysis workflow regarding drift correction and channel alignment has been copied from the cingulin studies to mimic these conditions.

Fluorescent beads (TetraSpeck™ Microspheres, 100 nm diameter, Thermo Fisher Scientific) that are covered with multiple fluorophore species and therefore emit under 488 nm, 561 nm as well as 647 nm excitation, are immobilized on a glass coverslip. They are imaged with the 100x oil objective and the 1.5x tube lens for 1000 frames in three channels consecutively (as performed for the cingulin studies) with maximum excitation power (without saturating the EMCCD camera). After imaging, the following analysis procedure is applied:

1. ThunderSTORM is used to detect the localizations for each color i.e. channel. By overlaying the three colors, separated and smeared out clusters can be seen throughout the whole field of view (Figure 9.7A, top panel), caused by continuous drift and possible chromatic aberrations.
2. A cross-correlation-based drift correction is applied per color (as for all the super-resolution studies in this work) which restores the circular shape of the clusters (Figure 9.7A, middle panel).
3. Due to the sequential imaging of the individual colors and therefore temporal separation, the three clusters from one bead are still separated and are therefore aligned using a rigid body transformation (MATLAB routine `pcregrigid`). Without any chromatic aberrations, this transformation should suffice to align the colors completely. As can be seen in the bottom panel of Figure 9.7A, this is not the case which shows that chromatic aberrations are present. In order to visualize the aberrations in more detail, Figure 9.7B displays clusters from each corner as well as the middle of the complete field of view (covering roughly 50 x 50 μm^2).

To quantify the offset between the colors, the cluster analysis routine as applied in chapter 3 for the DNA origami has been modified and programmed as follows:

4. The localizations from all three colors are pooled together and clustered using the established workflow from chapter 3.

5. Only clusters that contain at least 100 localizations of each color are kept.
6. Distances and vectors between the colors from a single bead are calculated by fitting each cluster with a single 2D Gaussian and comparing the peak positions. These distances are also displayed in Figure 9.7B as lines.
7. The channel separation between two colors can be visualized using a quiver plot. Figure 9.7C shows the offset between the 561 nm and 647 nm excitation channel. An increased offset for clusters further away from the center of the camera is detected, as is expected from chromatic aberrations.
8. By plotting the distance from the center against the offset for all clusters (Figure 9.7D), a linear dependence can be observed.

For the measured optical system, the chromatic aberrations between the 561 nm and 647 nm excitation channel after drift correction and channel alignment span from 0 nm in the center of the field of view up to around 100 nm in the corners. Similar behavior could be observed for the other channel combinations. In order to reduce chromatic aberrations in multicolor super-resolution images without a specialized correction [123], imaging in a cropped center region is preferred, as also applied for the cingulin studies in this work.

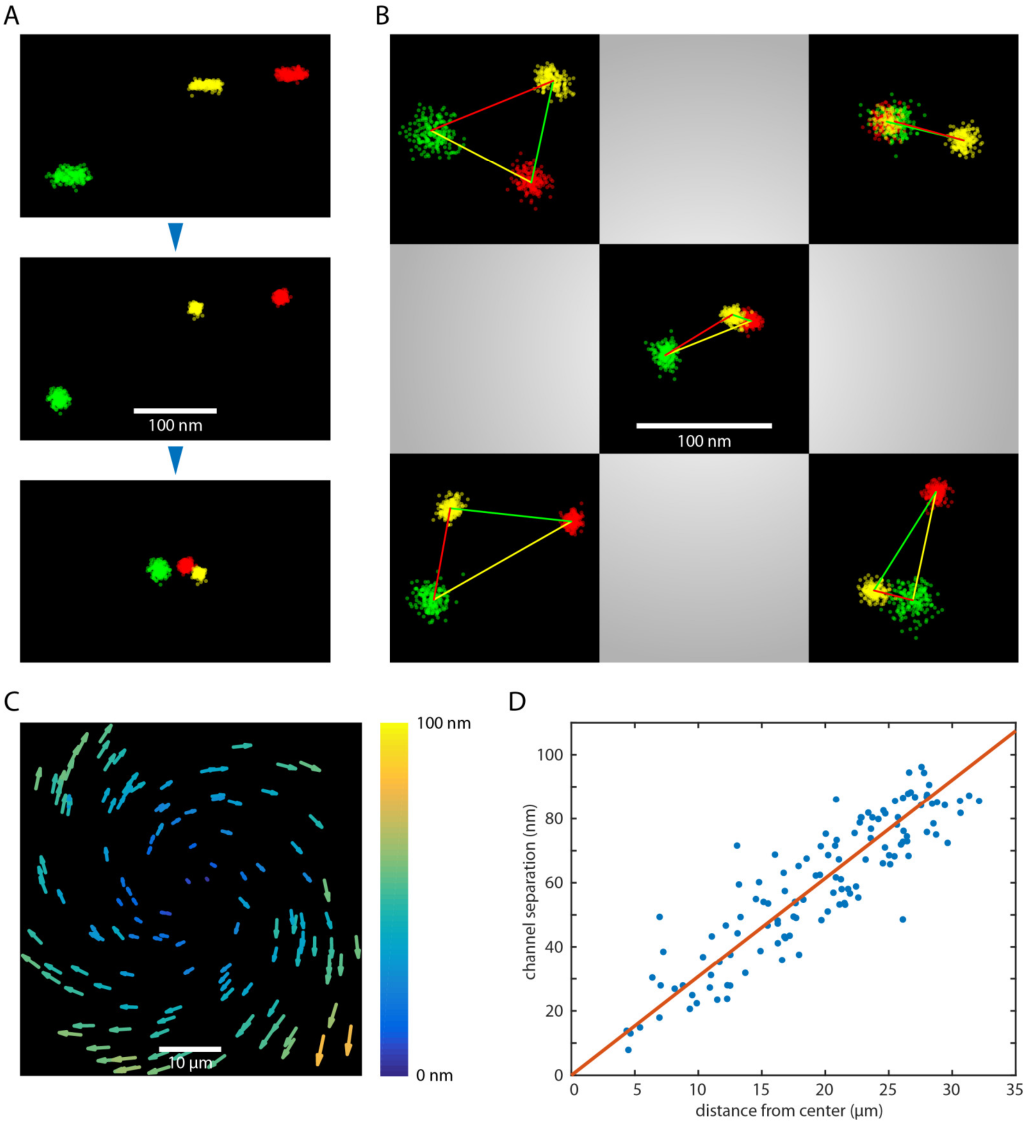


Figure 9.7 Quantifying chromatic aberrations with fluorescent beads A) Localizations from a single bead imaged in three channels (green: 488 nm, yellow: 561 nm, red: 647 nm excitation) are shown before drift-correction (top panel), after drift-correction (middle panel), and after channel alignment (bottom panel). B) Beads from the four corners and the middle of the camera's field of view are shown. Center positions and distances between the clusters are highlighted with lines. C) Quiver plot displaying the vectors from the red to the yellow peak position of individual beads (representing the channel separation - as displayed with green lines in plot B). Vector are colored by their lengths and have been scaled uniformly to enhance visibility. The whole camera field of view from a single measurement is shown. D) The vector length from plot C is plotted against the distance from the image center. A linear behavior is observed and fitted.

10 REFERENCES

[1] D. G. Mann and S. J. M. Droop, "3. Biodiversity, biogeography and conservation of diatoms," *Hydrobiologia*, vol. 336, no. 1–3, pp. 19–32, Oct. 1996.

[2] D. G. Mann, "The species concept in diatoms," *Phycologia*, vol. 38, no. 6, pp. 437–495, Nov. 1999.

[3] P. G. Falkowski, R. T. Barber, and V. Smetacek, "Biogeochemical controls and feedbacks on ocean primary production," *Science (80-.).*, vol. 281, no. 1998, pp. 200–206, 1998.

[4] B. A. Whitton, *Ecology of Cyanobacteria II*. Dordrecht: Springer Netherlands, 2012.

[5] N. Poulsen, P. M. Chesley, and N. Kröger, "Molecular Genetic Manipulation of the Diatom Thalassiosira Pseudonana (Bacillariophyceae)," *J. Phycol.*, vol. 42, no. 5, pp. 1059–1065, Oct. 2006.

[6] E. V. Armbrust *et al.*, "The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism.," *Science*, vol. 306, no. 5693, pp. 79–86, Oct. 2004.

[7] C. Bowler *et al.*, "The Phaeodactylum genome reveals the evolutionary history of diatom genomes.," *Nature*, vol. 456, no. 7219, pp. 239–44, 2008.

[8] Haeckel Ernst, *Kunstformen der Natur.*, vol. 1. Verlags des Bibliographischen Instituts in Leipzig und Wien, 1904.

[9] F. E. Round, R. M. Crawford, and D. G. Mann, *Diatoms: Biology and Morphology of the Genera*, 1st ed. Cambridge: Cambridge University Press, 2007.

[10] M. Hildebrand, "Biological processing of nanostructured silica in diatoms," *Prog. Org. Coatings*, vol. 47, pp. 256–266, 2003.

[11] N. Kröger and N. Poulsen, "Diatoms-from cell wall biogenesis to nanotechnology," *Annu. Rev. Genet.*, vol. 42, pp. 83–107, Jan. 2008.

[12] R. W. Drum and H. S. Pankratz, "Post mitotic fine structure of Gomphonema parvulum," *J. Ultrastruct. Res.*, vol. 10, no. 3, pp. 217–223, 1964.

[13] N. Kröger, R. Deutzmann, and M. Sumper, "Polycationic peptides from diatom biosilica that direct silica nanosphere formation.," *Science*, vol. 286, no. 5442, pp. 1129–1132, 1999.

[14] N. Poulsen and N. Kröger, "Silica morphogenesis by alternative processing of silaffins in the

diatom Thalassiosira pseudonana," *J. Biol. Chem.*, vol. 279, no. 41, pp. 42993–42999, 2004.

[15] S. Wenzl, R. Hett, P. Richthammer, and M. Sumper, "Silacidins: Highly acidic phosphopeptides from diatom shells assist in silica precipitation in vitro," *Angew. Chemie Int. Ed.*, vol. 47, no. 9, pp. 1729–1732, 2008.

[16] M. Sumper, E. Brunner, and G. Lehmann, "Biomineralization in diatoms: Characterization of novel polyamines associated with silica," *FEBS Lett.*, vol. 579, no. 17, pp. 3765–3769, 2005.

[17] A. Scheffel, N. Poulsen, S. Shian, and N. Kröger, "Nanopatterned protein microrings from a diatom that direct silica morphogenesis.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 8, pp. 3175–3180, Feb. 2011.

[18] A. Kotzsch *et al.*, "Biochemical Composition and Assembly of Biosilica-associated Insoluble Organic Matrices from the Diatom Thalassiosira pseudonana.," *J. Biol. Chem.*, vol. 291, no. 10, pp. 4982–97, Mar. 2016.

[19] E. Brunner *et al.*, "Chitin-based organic networks: an integral part of cell wall biosilica in the diatom Thalassiosira pseudonana.," *Angew. Chemie Int. Ed.*, vol. 48, no. 51, pp. 9724–7, 2009.

[20] N. Poulsen, A. Scheffel, V. C. Sheppard, P. M. Chesley, and N. Kröger, "Pentalysine clusters mediate silica targeting of silaffins in Thalassiosira pseudonana.," *J. Biol. Chem.*, vol. 288, no. 28, pp. 20100–9, Jul. 2013.

[21] N. Poulsen, C. Berne, J. Spain, and N. Kröger, "Silica immobilization of an enzyme through genetic engineering of the diatom Thalassiosira pseudonana.," *Angew. Chemie Int. Ed.*, vol. 46, no. 11, pp. 1843–1846, Jan. 2007.

[22] J. Pickett-Heaps, A.-M. M. Schmid, and L. A. Edgar, "The cell biology of diatom valve formation," *Prog. Phycol. Res.*, vol. 7, pp. 1–168, 1990.

[23] M. Hildebrand *et al.*, "Nanoscale control of silica morphology and three-dimensional structure during diatom cell wall formation," *J. Mater. Res.*, vol. 21, no. 10, pp. 2689–2698, Oct. 2006.

[24] F. Wolfe-Simon, V. Starovoytov, J. R. Reinfelder, O. Schofield, and P. G. Falkowski, "Localization and role of manganese superoxide dismutase in a marine diatom.," *Plant Physiol.*, vol. 142, no. 4, pp. 1701–9, Dec. 2006.

[25] B. J. Karas *et al.*, "Designer diatom episomes delivered by bacterial conjugation.," *Nat. Commun.*, vol. 6, p. 6925, Apr. 2015.

[26] A. Hopes, V. Nekrasov, S. Kamoun, and T. Mock, "Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana.," *Plant Methods*, vol. 12, p. 49, 2016.

[27] M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. SHORT COMMUNICATION," *J. Microsc.*, vol. 198, no. 2, pp. 82–87, May 2000.

[28] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.*, vol. 19, no. 11, p. 780, Jun. 1994.

[29] D. Axelrod, "Cell-substrate contacts illuminated by total internal reflection fluorescence.," *J. Cell Biol.*, vol. 89, no. 1, pp. 141–5, Apr. 1981.

[30] L. S. Barak and W. W. Webb, "Fluorescent low density lipoprotein for observation of dynamics of individual receptor complexes on cultured human fibroblasts.," *J. Cell Biol.*, vol. 90, no. 3, pp. 595–604, Sep. 1981.

[31] E. Betzig *et al.*, "Imaging intracellular fluorescent proteins at nanometer resolution.," *Science*, vol. 313, no. 5793, pp. 1642–1645, Sep. 2006.

[32] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by

fluorescence photoactivation localization microscopy.," *Biophys. J.*, vol. 91, no. 11, pp. 4258–72, Dec. 2006.

[33] M. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods*, vol. 3, no. 10, pp. 793–795, 2006.

[34] J. Lippincott-Schwartz and G. H. Patterson, "Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging.," *Trends Cell Biol.*, vol. 19, no. 11, pp. 555–65, Nov. 2009.

[35] D. Bourgeois and V. Adam, "Reversible photoswitching in fluorescent proteins: A mechanistic view," *IUBMB Life*, vol. 64, no. June, pp. 482–491, 2012.

[36] J. Vogelsang *et al.*, "Make them blink: probes for super-resolution microscopy.," *Chemphyschem*, vol. 11, no. 12, pp. 2475–90, Aug. 2010.

[37] S. van de Linde and M. Sauer, "How to switch a fluorophore: from undesired blinking to controlled photoswitching," *Chem. Soc. Rev.*, vol. 43, no. 4, pp. 1076–1087, 2014.

[38] G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, and X. Zhuang, "Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging.," *Nat. Methods*, vol. 8, no. 12, pp. 1027–36, Dec. 2011.

[39] P. Hoyer, T. Staudt, J. Engelhardt, and S. W. Hell, "Quantum dot blueing and blinking enables fluorescence nanoscopy.," *Nano Lett.*, vol. 11, no. 1, pp. 245–50, Jan. 2011.

[40] E. H. Chen, O. Gaathon, M. E. Trusheim, and D. Englund, "Wide-Field Multispectral Super-Resolution Imaging Using Spin-Dependent Fluorescence in Nanodiamonds," *Nano Lett.*, vol. 13, no. 5, pp. 2073–2077, May 2013.

[41] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-Dimensional Super-Resolution Reconstruction Microscopy," *Science (80-.).*, vol. 319, no. February, pp. 810–813, 2008.

[42] A. R. Carter, G. M. King, T. A. Ulrich, W. Halsey, D. Alchenberger, and T. T. Perkins, "Stabilization of an optical microscope to 0.1 nm in three dimensions.," *Appl. Opt.*, vol. 46, no. 3, pp. 421–7, Jan. 2007.

[43] Y. Wang *et al.*, "Localization events-based sample drift correction for localization microscopy with redundant cross-correlation algorithm.," *Opt. Express*, vol. 22, no. 13, pp. 15982–91, Jun. 2014.

[44] H. Shroff, C. G. Galbraith, J. A. Galbraith, and E. Betzig, "Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics," *Nat. Methods*, vol. 5, no. 5, pp. 417–423, 2008.

[45] P. Almada, S. Culley, and R. Henriques, "PALM and STORM: Into large fields and high-throughput microscopy with sCMOS detectors.," *Methods*, vol. 88, pp. 109–21, Oct. 2015.

[46] K. Kwakwa *et al.*, "easySTORM: a robust, lower-cost approach to localisation and TIRF microscopy.," *J. Biophotonics*, vol. 9, no. 9, pp. 948–57, Sep. 2016.

[47] D. Sage *et al.*, "Quantitative evaluation of software packages for single-molecule localization microscopy.," *Nat. Methods*, vol. 12, no. 8, pp. 717–24, Aug. 2015.

[48] J. J. Schmied *et al.*, "DNA origami-based standards for quantitative fluorescence microscopy.," *Nat. Protoc.*, vol. 9, no. 6, pp. 1367–91, Jan. 2014.

[49] M. Ovesný, P. Křížek, J. Borkovec, Z. Svindrych, and G. M. Hagen, "ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging.," *Bioinformatics*, vol. 30, no. 16, pp. 2389–2390, Aug. 2014.

[50] I. Izeddin *et al.*, "Wavelet analysis for single molecule localization microscopy.," *Opt. Express*, vol. 20, no. 3, pp. 2081–95, Jan. 2012.

[51] M. J. Mlodzianoski *et al.*, "Sample drift correction in 3D fluorescence photoactivation

localization microscopy.," *Opt. Express*, vol. 19, no. 16, pp. 15009–19, Aug. 2011.

[52] D. Baddeley, M. B. Cannell, and C. Soeller, "Visualization of localization microscopy data.," *Microsc. Microanal.*, vol. 16, no. 1, pp. 64–72, Feb. 2010.

[53] R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescent probes.," *Biophys. J.*, vol. 82, no. 5, pp. 2775–2783, May 2002.

[54] C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[55] N. Banterle, K. H. Bui, E. A. Lemke, and M. Beck, "Fourier ring correlation as a resolution criterion for super-resolution microscopy," *J. Struct. Biol.*, vol. 183, no. 3, pp. 363–367, 2013.

[56] R. P. J. Nieuwenhuizen *et al.*, "Measuring image resolution in optical nanoscopy.," *Nat. Methods*, vol. 10, no. 6, pp. 557–562, Jun. 2013.

[57] F. Levet *et al.*, "SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data," *Nat. Methods*, vol. 12, no. 11, 2015.

[58] L. Andronov, I. Orlov, Y. Lutz, J. Vonesch, and B. P. Klaholz, "ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy.," *Sci. Rep.*, vol. 6, no. April, p. 24084, Apr. 2016.

[59] B. D. Ripley, "Tests of `Randomness' for Spatial Point Patterns," *J. R. Stat. Soc. Ser. B*, vol. 41, no. 3, pp. 368–374, 1979.

[60] M. A. Kiskowski, J. F. Hancock, and A. K. Kenworthy, "On the use of Ripley's K-function and its derivatives to analyze domain size.," *Biophys. J.*, vol. 97, no. 4, pp. 1095–103, Aug. 2009.

[61] X. Ester, M., Kriegel, H.-P., Sander, J., Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proc. 2nd Int. Conf. Knowl. Discov. Data Min.*, pp. 226–231, 1996.

[62] J. Macqueen, "Some methods for classification and analysis of multivariate observations," *Proc. Fifth Berkeley Symp. Math. Stat. Probab.*, vol. 1, no. 233, pp. 281–297, 1967.

[63] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.

[64] D. Arthur and S. Vassilvitskii, "K-Means++: the Advantages of Careful Seeding," *Proc. eighteenth Annu. ACM-SIAM Symp. Discret. algorithms*, vol. 8, pp. 1027–1025, 2007.

[65] K. W. Dunn, M. M. Kamocka, and J. H. McDonald, "A practical guide to evaluating colocalization in biological microscopy.," *Am. J. Physiol. Cell Physiol.*, vol. 300, no. 4, pp. C723-42, Apr. 2011.

[66] S. Malkusch, U. Endesfelder, J. Mondry, M. Gelléri, P. J. Verveer, and M. Heilemann, "Coordinate-based colocalization analysis of single-molecule localization microscopy data," *Histochem. Cell Biol.*, vol. 137, no. 1, pp. 1–10, 2012.

[67] M. Georgieva, D. I. Cattoni, J.-B. Fiche, T. Mutin, D. Chamousset, and M. Nollmann, "Nanometer resolved single-molecule colocalization of nuclear factors by two-color super resolution microscopy imaging," *Methods*, 2016.

[68] S. Wolter, A. Löschberger, and T. Holm, "rapidSTORM: accurate, fast open-source software for localization microscopy," *Nat. ...*, vol. 9, no. 11, pp. 1040–1042, 2012.

[69] R. Henriques, M. Lelek, E. F. Fornasiero, F. Valtorta, C. Zimmer, and M. M. Mhlanga, "QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ.," *Nat. Methods*, vol. 7, no. 5, pp. 339–40, May 2010.

[70] Y. Tang, X. Wang, X. Zhang, J. Li, and L. Dai, "Sub-nanometer drift correction for super-resolution imaging.," *Opt. Lett.*, vol. 39, no. 19, pp. 5685–8, Oct. 2014.

[71] P. Bon *et al.*, “Three-dimensional nanometre localization of nanoparticles to enhance super-resolution microscopy,” *Nat. Commun.*, vol. 6, p. 7764, 2015.

[72] P. W. K. Rothemund, “Folding DNA to create nanoscale shapes and patterns,” *Nature*, vol. 440, no. 7082, pp. 297–302, 2006.

[73] A. Keller, I. Bald, A. Rotaru, E. Cauet, K. V Gothelf, and F. Besenbacher, “Probing Electron-Induced Bond Cleavage at the Single-Molecule Level Using DNA Origami Templates,” *ACS Nano*, vol. 6, no. 5, pp. 4392–4399, 2012.

[74] K. Nienhaus and G. U. Nienhaus, “Fluorescent proteins for live-cell imaging with super-resolution.,” *Chem. Soc. Rev.*, vol. 43, no. 4, pp. 1088–106, Feb. 2014.

[75] Y. Ishitsuka, K. Nienhaus, and G. U. Nienhaus, “Photoactivatable fluorescent proteins for super-resolution microscopy.,” *Methods Mol. Biol.*, vol. 1148, pp. 239–60, 2014.

[76] D. M. Shcherbakova, P. Sengupta, J. Lippincott-Schwartz, and V. V Verkhusha, “Photocontrollable Fluorescent Proteins for Superresolution Imaging.,” *Annu. Rev. Biophys.*, vol. 43, pp. 303–329, May 2014.

[77] J. Wiedenmann *et al.*, “EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 45, pp. 15905–10, Nov. 2004.

[78] R. Ando, H. Hama, M. Yamamoto-Hino, H. Mizuno, and A. Miyawaki, “An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 20, pp. 12651–6, Oct. 2002.

[79] D. M. Chudakov *et al.*, “Kindling fluorescent proteins for precise in vivo photolabeling.,” *Nat. Biotechnol.*, vol. 21, no. 2, pp. 191–194, Feb. 2003.

[80] V. Adam *et al.*, “Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 47, pp. 18343–18348, 2008.

[81] G. H. Patterson and J. Lippincott-Schwartz, “A photoactivatable GFP for selective photolabeling of proteins and cells.,” *Science*, vol. 297, no. 5588, pp. 1873–1877, Sep. 2002.

[82] F. V. Subach, G. H. Patterson, M. Renz, J. Lippincott-Schwartz, and V. V. Verkhusha, “Bright monomeric photoactivatable red fluorescent protein for two-color super-resolution sptPALM of live cells,” *J. Am. Chem. Soc.*, vol. 132, no. 18, pp. 6481–6491, 2010.

[83] F. V Subach, G. H. Patterson, S. Manley, J. M. Gillette, J. Lippincott-Schwartz, and V. V Verkhusha, “Photoactivatable mCherry for high-resolution two-color fluorescence microscopy.,” *Nat. Methods*, vol. 6, no. 2, pp. 153–9, Feb. 2009.

[84] N. G. Gurskaya *et al.*, “Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light,” *Nat. Biotechnol.*, vol. 24, no. 4, pp. 461–465, 2006.

[85] M. Zhang *et al.*, “Rational design of true monomeric and bright photoactivatable fluorescent proteins.,” *Nat. Methods*, vol. 9, no. 7, pp. 727–9, May 2012.

[86] R. Ando, H. Mizuno, and A. Miyawaki, “Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting.,” *Science*, vol. 306, no. 5700, pp. 1370–3, Nov. 2004.

[87] D. M. Chudakov, V. V Verkhusha, D. B. Staroverov, E. A. Souslova, S. Lukyanov, and K. A. Lukyanov, “Photoswitchable cyan fluorescent protein for protein tracking,” *Nat. Biotechnol.*, vol. 22, no. 11, pp. 1435–1439, 2004.

[88] O. M. Subach, G. H. Patterson, L.-M. Ting, Y. Wang, J. S. Condeelis, and V. V Verkhusha, “A photoswitchable orange-to-far-red fluorescent protein, PSmOrange.,” *Nat. Methods*, vol. 8, no. 9, pp. 771–7, Jul. 2011.

[89] S. Wang, J. R. Moffitt, G. T. Dempsey, X. S. Xie, and X. Zhuang, “Characterization and

development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 23, pp. 8452–7, Jun. 2014.

[90] M. Andresen *et al.*, "Photoswitchable fluorescent proteins enable monochromatic multilabel imaging and dual color fluorescence nanoscopy.," *Nat. Biotechnol.*, vol. 26, no. 9, pp. 1035–40, Sep. 2008.

[91] D. K. Tiwari *et al.*, "A fast- and positively photoswitchable fluorescent protein for ultralow-laser-power RESOLFT nanoscopy.," *Nat. Methods*, vol. 12, no. 6, pp. 515–8, Jun. 2015.

[92] A. C. Stiel *et al.*, "Generation of monomeric reversibly switchable red fluorescent proteins for far-field fluorescence nanoscopy.," *Biophys. J.*, vol. 95, no. 6, pp. 2989–97, Sep. 2008.

[93] T. Grotjohann *et al.*, "Diffraction-unlimited all-optical imaging and writing with a photochromic GFP.," *Nature*, vol. 478, no. 7368, pp. 204–8, Sep. 2011.

[94] P. Kuczynska, M. Jemiola-Rzeminska, and K. Strzalka, "Photosynthetic Pigments in Diatoms.," *Mar. Drugs*, vol. 13, no. 9, pp. 5847–81, Sep. 2015.

[95] F. Sievers *et al.*, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.," *Mol. Syst. Biol.*, vol. 7, no. 1, p. 539, 2011.

[96] N. P. Lemay, A. L. Morgan, E. J. Archer, L. A. Dickson, C. M. Megley, and M. Zimmer, "The Role of the Tight-Turn, Broken Hydrogen Bonding, Glu222 and Arg96 in the Post-translational Green Fluorescent Protein Chromophore Formation.," *Chem. Phys.*, vol. 348, no. 1–3, pp. 152–160, Jun. 2008.

[97] J. A. J. Arpino, P. J. Rizkallah, and D. D. Jones, "Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222.," *PLoS One*, vol. 7, no. 10, p. e47132, 2012.

[98] J. J. van Thor, T. Gensch, K. J. Hellingwerf, and L. N. Johnson, "Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222.," *Nat. Struct. Biol.*, vol. 9, no. 1, pp. 37–41, 2002.

[99] F. V. Subach and V. V. Verkhusha, "Chromophore transformations in red fluorescent proteins," *Chem. Rev.*, vol. 112, pp. 4308–4327, 2012.

[100] F. V Subach *et al.*, "Photoactivation mechanism of PAmCherry based on crystal structures of the protein in the dark and fluorescent states.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 50, pp. 21097–21102, 2009.

[101] H. Mizuno *et al.*, "Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein," *Mol. Cell*, vol. 12, no. 4, pp. 1051–1058, 2003.

[102] M. Andresen *et al.*, "Structural basis for reversible photoswitching in Dronpa.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 32, pp. 13005–135009, Aug. 2007.

[103] K. Nienhaus and G. U. Nienhaus, "Chromophore photophysics and dynamics in fluorescent proteins of the GFP family," *J. Phys. Condens. Matter*, vol. 28, no. 44, p. 443001, 2016.

[104] T. J. Chozinski, L. a Gagnon, and J. C. Vaughan, "Twinkle, Twinkle Little Star: Photoswitchable Fluorophores for Super-Resolution Imaging.," *FEBS Lett.*, no. July, Jul. 2014.

[105] M. Hildebrand, S. Kim, D. Shi, K. Scott, and S. Subramaniam, "3D imaging of diatoms with ion-abrasion scanning electron microscopy.," *J. Struct. Biol.*, vol. 166, no. 3, pp. 316–328, Jun. 2009.

[106] X. Chen, C. Wang, E. Baker, and C. Sun, "Numerical and experimental investigation of light trapping effect of nanostructured diatom frustules," *Sci. Rep.*, vol. 5, no. January, p. 11977, 2015.

[107] J. Romann, J.-C. Valmalette, M. S. Chauton, G. Tranell, M.-A. Einarsrud, and O. Vadstein, “Wavelength and orientation dependent capture of light by diatom frustule nanostructures.,” *Sci. Rep.*, vol. 5, no. 7491, p. 17403, 2015.

[108] J. Li, D. M. Czajkowsky, X. Li, and Z. Shao, “Fast immuno-labeling by electrophoretically driven infiltration for intact tissue imaging,” *Sci. Rep.*, vol. 5, no. April, p. 10640, 2015.

[109] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, “On the shape of a set of points in the plane,” *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 551–559, Jul. 1983.

[110] A. Monserrate, S. Casado, and C. Flors, “Correlative atomic force microscopy and localization-based super-resolution microscopy: revealing labelling and image reconstruction artefacts.,” *Chemphyschem*, vol. 15, no. 4, pp. 647–50, Mar. 2014.

[111] M. Perkovic *et al.*, “Correlative Light- and Electron Microscopy with chemical tags,” *J. Struct. Biol.*, vol. 186, no. 2, pp. 205–213, 2014.

[112] A. Loschberger, C. Franke, G. Krohne, S. van de Linde, and M. Sauer, “Correlative super-resolution fluorescence and electron microscopy of the nuclear pore complex with molecular resolution,” *J. Cell Sci.*, vol. 127, no. 20, pp. 4351–4355, 2014.

[113] N. R. Ford *et al.*, “Antigen Binding and Site-Directed Labeling of Biosilica-Immobilized Fusion Proteins Expressed in Diatoms.,” *ACS Synth. Biol.*, vol. 5, no. 3, pp. 193–199, Mar. 2016.

[114] T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein, “Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI).,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 52, pp. 22287–22292, Dec. 2009.

[115] S.-H. Shim *et al.*, “Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 35, pp. 13978–83, Aug. 2012.

[116] P. J. Harrison, R. E. Waters, and F. J. R. Taylor, “A Broad Spectrum Artificial Sea Water Medium for Coastal and Open Ocean Phytoplankton1,” *J. Phycol.*, vol. 16, no. 1, pp. 28–35, Mar. 1980.

[117] N. Kroger, “Self-Assembly of Highly Phosphorylated Silaffins and Their Function in Biosilica Morphogenesis,” *Science (80-.).*, vol. 298, no. 5593, pp. 584–586, Oct. 2002.

[118] R. Roy, S. Hohng, and T. Ha, “A practical guide to single-molecule FRET.,” *Nat. Methods*, vol. 5, no. 6, pp. 507–16, Jun. 2008.

[119] A. Kotzsch *et al.*, “Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization,” *Manuscipt submitted*, 2017.

[120] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, “Efficient subpixel image registration algorithms.,” *Opt. Lett.*, vol. 33, no. 2, pp. 156–8, Jan. 2008.

[121] S. Geissbuehler, N. L. Bocchio, C. Dellagiacoma, C. Berclaz, M. Leutenegger, and T. Lasser, “Mapping molecular statistics with balanced super-resolution optical fluctuation imaging (bSOFI),” *Opt. Nanoscopy*, vol. 1, no. 1, p. 4, 2012.

[122] S. C. Stein, A. Huss, D. Hähnel, I. Gregor, and J. Enderlein, “Fourier interpolation stochastic optical fluctuation imaging,” *Opt. Express*, vol. 23, no. 12, p. 16154, Jun. 2015.

[123] M. Erdelyi *et al.*, “Correcting chromatic offset in multicolor super-resolution localization microscopy,” *Opt. Express*, vol. 21, no. 9, pp. 10978–10988, 2013.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Prof. Dr. Michael Schlierf for letting me work on this PhD project in his group. His guidance, support, and positive attitude enabled me to successful finish this thesis. Additionally, I greatly enjoyed the possibility to explore additional projects that allowed me to expand my skillset, support other group members and led to fruitful collaborations – all of which was supported by Michael.

I enjoyed my time in the Schlierf-Lab very much and have to thank all their members that added to a very friendly working atmosphere and made the last years very enjoyable. Many colleagues have become close friends of mine since the start of my thesis. In particular, I want to thank Dr. Christiane Jungnickel and Andreas Hartmann for their constant scientific and moral support. Additionally, Ashwin Balakrishnan helped me greatly in DNA origami measurements. I thank Kristin Wogan and Kristin Hunold for supporting me in wet lab work, and Katja Dornig for sorting out all the bureaucratic obstacles on my way. Thanks to all of you!

Furthermore, I have to thank various members of the Kröger-Lab for their help in diatom-related work. Prof. Dr. Nils Kröger and Dr. Nicole Poulsen supported me greatly throughout the doctoral work and have been patient in trying to teach a physicist diatom biology. I thank Damian Pawolski for his help in SEM imaging, general consultation, and his dry humor. I am grateful for Jennifer Klemm's and Anke Borrmann's unweary support in performing diatom transformations for me.

I would like to thank the members of my thesis committee for their time: Prof. Dr. Michael Schlierf, Prof. Dr. Nils Kröger, Prof. Dr. Thorsten Mascher, and in particular Prof. Dr. Stefan Diez who agreed on reviewing my thesis as well.

Last but not least, I want to thank some people on a more personal note. My parents have supported me throughout all the years to continue no matter what I was doing. They helped me whenever I needed support … or just a delicious meal. I am very fortunate to have you as my parents. Finally, I will thank Lisa for her tireless assistance in correcting my thesis. Outside science, I have to thank Lisa for her daily support and love and for making the last years the best of my life. I am very happy to have you in my life and looking forward to our new family!

No matter how politely one says it,
we owe our existence to the farts of blue-green algae.

- Diane Ackerman

DECLARATION

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die Dissertation wurde im Zeitraum von Mai 2014 bis Mai 2017 verfasst und von Prof. Dr. Michael Schlierf, B CUBE - Center for Molecular Bioengineering, Technische Universität Dresden, betreut.

Meine Person betreffend erkläre ich hiermit, dass keine früheren erfolglosen Promotionsverfahren stattgefunden haben.

Ich erkenne die Promotionsordnung der Fakultät für Mathematik und Naturwissenschaften, Technische Universität Dresden an.

Datum, Ort	Unterschrift